

2 December 1982

Where not to put strategic missiles

The Reagan-Weinberger solution to the MX-siting problem is ingenious, probably unworkable and certain to confuse the generals. It points only towards the need for strategic arms control.

The Reagan Administration appears to be capable of thinking not merely the unthinkable but the unworkable. Now, after years of uncertainty, Mr Caspar Weinberger, the US Secretary of Defense, has proposed and President Ronald Reagan has agreed that 100 MX missiles should be based in a narrow tract of land in Wyoming, handy for the short journey over the North Pole to targets in the Soviet Union but also so well advertised in position as to be itself an easy target for Soviet intercontinental missiles. In the bewildering spectrum of solutions to the problem of how the MX missiles should be sited, the Reagan-Weinberger solution is at the other extreme from that to which President Carter appeared to be most attracted — building underground railway systems so that 100 missiles could be distributed more or less at random (so far as outsiders could tell) between more than 2,000 potential launching silos. The Carter solution to the problem of where and how to put the missiles had the obvious advantage that a potential enemy of the United States (supposed for these purposes to be the Soviet Union) could make an effective first strike against the US stock of strategic missiles only by making 2,000 separate attacks on as many holes in the ground, of which fewer than five per cent would be occupied by missiles. The disadvantage, as the people of the western states told Mr Carter clearly enough, was that the whole arrangement would be an infernal nuisance, socially as well as environmentally. His taxpayers were also understandably alarmed at the cost. The Reagan-Weinberger solution has the advantage of avoiding the most obvious difficulties. Its disadvantage, which influential people as different as Professor Charles Townes (see *Nature* 18 November, p.208) and Dr Richard Garwin are pointing out, is that it lacks credibility.

Although intercontinental ballistic missiles are only a little more than a quarter of a century old, they have been with us long enough for the question of where to put them to be recognized as important, crucial and vital (in the sense of life and death). At the outset, when missiles were so inaccurate that their only accessible targets had to be large cities, and when the chance of a missile silo being hit by an attacking missile aimed at it was virtually zero, superpowers could base their intercontinental missiles in silos whose locations were known to potential enemies without fearing that their capacity to retaliate would be impaired by a preemptive strike. For the past decade, however, and especially with the development of techniques for guiding warheads to their targets without continual reference to the launching site (i.e. intelligently), that calculation has gone by the board. The chances now are that it is possible for one side or the other to destroy point targets on the other side with weapons whose yield in terms of the equivalent in conventional explosives is no more than, say, 100,000 tonnes. So the United States is rightly worried that its force of nearly 1,000 Minutemen scattered through the western states has become vulnerable to a first strike, and anxious that it should be replaced by something better or less vulnerable. Even those who disdain entirely the basis on which such calculations are made must acknowledge that a superpower worried about the viability of its striking force is a dangerous superpower, likely to be tempted to fire them off before there is an undeniable need to do so. In other words, so long as superpowers exist, they must be helped to feel confident about the security of their missile forces for otherwise they will become trigger-happy and dangers to the rest of us.

The principal Soviet striking force is no better placed. While missile launchers are able to shuttle along most of the length of the trans-Siberian railway and thus secure a measure of protection from uncertainty, the speed with which they move (slow) and the steady improvement of satellite reconnaissance also makes them vulnerable to a first strike from some other superpower. So there is a case for hoping that both of those superpowers with a legitimate interest in the maintenance of counter-force missile flights should be helped towards a system for their deployment that will be more credible for the generals, more conducive to good nights' sleep by generals' political bosses (and less burdensome for taxpayers). What can this be?

The Reagan-Weinberger solution in Wyoming is a reasonable starting point. The calculation is that a hostile superpower would first attack the southernmost missile silos; otherwise succeeding waves of missiles would necessarily have to travel through electrically and mechanically disturbed atmosphere. So why not build the southernmost silos, but not put missiles in them? With a little luck, it might be possible to get by with only half the force of 100 missiles now planned. But with a little ingenuity, it will be possible to save still more public money and at the same time to reinforce the now shaken confidence of the generals, by leaving all the silos empty but deciding, secretly, to put a few MX missiles in existing holes in the ground, the Minuteman silos soon to be vacated for example. Or it could, with an even greater exercise of ingenuity and intelligence, be recognized by both the superpowers that what the Reagan-Weinberger solution of the MX problem demonstrates is the need for both of them to opt out of the application of a first strike capability.

In retrospect, the 1950s, when the two superpowers were able only to threaten each other with the crudest forms of annihilation, were comparatively comforting times. The gulf between conventional conflict and mutual nuclear annihilation was huge — so huge that nobody ever thought it would be bridged. What technology has made possible, in the MX missile and its Soviet counterparts, is the prospect that the other fellow's strategic force might be destroyed before it can be used, coupled with a haunting anxiety that the missiles that will accomplish that cannot themselves be hidden or protected. The sensible course, therefore, is to agree to give them up, by negotiation at Geneva. For the time being, neither superpower need abandon its force of submarine-launched missiles, the reason why each can say to the other that annihilation of people, not machines, is still within its grasp. But the Reagan-Weinberger solution should, by the fancifulness of its disposition of expensive hardware, point urgently in that direction.

The Reagan-Weinberger solution of this dilemma is ingenious and even daring but may not create the sense of confidence that both generals and taxpayers in the United States are likely to demand. The theory, of course, is simple. If a hostile warhead should attack some missile silo in the proposed "Dense Pack" missile park in Wyoming, destroying that and putting its immediate neighbours out of action, the calculation is that there would still be others (further north?) still capable of reacting, and that these would be invulnerable to attacking warheads because they would be protected by the atmospheric ionization (radio blackout) and turbulence caused by the first attacks. The underlying assumption, of course, is that the attacking warheads



would be fashioned as those intended to be packed (ten to a rocket) in the MX missile are designed. The trouble is, for the Reagan-Weinberger solution, that attacking warheads might be differently designed, made so as to make large holes in the ground. But in any case, what the generals in charge of the proposed missile park in Wyoming will know up until the last moment is that there has not and will not have been a practical demonstration of the doctrine of fratricide, the argument that the first attacking warhead will prevent others from doing their job. And in the event, uncertainty in a general's mind is more important than any amount of theoretical calculation by professional people who have studied the effects of nuclear weapons on the atmosphere. The calculations may be right, but if the generals do not accept their implications, they may ask the president of their day (unlikely to be Mr Ronald Reagan) to press the button that will prevent a preemptive attack. And he (or she) will have no choice but to agree.

Tell us about our friends

The President of the Royal Society has made a brave speech on Soviet illiberality.

How should a learned society, or for that matter a national rugby football union, conduct international relations with its counterparts in countries whose governments command little respect? This old dilemma is too little discussed, but Sir Andrew Huxley seems to have enlivened this year's anniversary meeting of the Royal Society (see next page) by an account of how deliberate and even obdurate innocence can triumph where diplomacy dictates mere compromise.

Conventional logic goes like this. By convention, (and United Nations charter) governments do not interfere in each others' internal affairs except in those mercifully rare circumstances in which they go to war with each other. National academies (not to mention rugby football unions) are even less well-equipped to effect political change in any sphere. Their analysis of what needs changing, and how, is certain to be less expert than that of the political professionals, diplomats on the international stage. Their assessment of how change might be accomplished is bound to be amateurish. Their resolution to carry through a recipe for change is bound to be compromised by differences of opinion among their members. This is why the Royal Society has been right, in this presidency and the last (Lord Todd's), to insist that the character of the Soviet government was not part of its concern, and was in any case a question outside its competence. If Soviet citizens are locked up or harassed in other ways because Soviet laws expect and exact more from them than do other people's laws, the fact that those injured happen to be scientists (or rugby footballers) is strictly speaking irrelevant. Indeed, when the going is rough, conventional opinion has it that the preservation of close relationships is all the more important. This is the spirit in which Huxley this week insisted that the pitiful trickle of professional exchanges with the Soviet Union should not be compromised by his decision to complain in public about the Soviet academy's apparently deliberate misunderstanding of questions about the ill-treatment of certain Soviet scientists.

Huxley's predecessor, Lord Todd, maddened much of the British scientific community by his proper repetition of the doctrine that sovereign governments of any kind are autonomous within their legal territory, but then won unexpected friends with his fierce attack on Soviet oppression of scientists at the Hamburg meeting on the Helsinki accords three years ago. Huxley's line has been more guileful. Enquire about your friends. When challenged, put the details down on paper. Then hope for the best but anticipate the worst. And when the weasel answer comes that Sakharov has been dealt with leniently, and when you know that he is still confined in Gork'i with the third winter closing in, simply say in public that the explanation is insufficient. The doctrine of the autonomy of sovereign states is an essential foundation of orderliness, and should not be undermined, but what is to prevent us asking after our friends?

Adam Smith forgotten

This week's meeting on international trade could have been disastrous but should have been better.

In the nine years since the last formal meeting of the governments that are party to the General Agreement on Tariffs and Trade (GATT), the world for which the agreement was originally negotiated has been transformed out of recognition. Nine years ago, the great increase of the international price of oil had only just begun, and few could guess what the consequences would be. Now, of course, we know: inflation and recession in the industrialized world, inflation and stagnation elsewhere. The rich countries are at best marking time, (but Japan, ironically classified at GATT as a country in the process of industrialization, is an exception); the poor are mostly becoming poorer (but countries such as India may yet be exceptions). The barely hidden agenda for this year's meeting seems to have been the wish of industrialized countries to look each other in the eye so as to discover whether they should hold to the principles of free trade on which GATT is founded (and if so whether their partners could be persuaded to obey the rules) or whether they should follow the politically more prudent course and settle for protectionism instead. Although Monday's formal communique reaffirms the principle that free trade is best, only in the months ahead will it be clear how many of the signatories believe that to be the case. (The government of Australia walked out at the weekend on the grounds that the then promised declaration was too leaky a commitment, generously studded as it was with phrases such as "wherever practicable".)

The trouble with GATT, like any other club, is that its members can and will resign whenever they think the costs of membership outweigh the benefits. And the trouble with a club whose members agree to trade freely with each other is that the costs are the more easily perceived. Most industrialized countries, now struggling with unpalatable and perhaps even unsupportable unemployment, know that measures to keep out imports would have an immediately beneficial effect. Domestic customers now buying efficiently-made goods imported from abroad would instead buy less efficiently manufactured domestic products, creating jobs for fellow-countrymen. Only in the longer run would the snags show up — customers required to support inefficient industries are on that account less prosperous than they might be. Two centuries ago, Adam Smith made the point quite simply: the wealth of nations stems from their capacity to bring about a rational division of labour, domestically and internationally.

This time round, GATT has merely stayed still. The constructive parts of the communique on Monday are those referring to studies to be carried out in the first instance by officials on some of the more contentious issues — should free trade extend not merely to manufactured goods but to invisibles — insurance, shipping and the like? (Yes, whatever Japan says.) Should free trade also extend to agriculture? (Yes, even though the European Community is outwardly united in its opposition.) What should be done about counterfeiting? (It should be penalized.) And the plight of countries in the process of industrialization? Who knows? If GATT hopes to survive, it must answer these questions before too many of its members drift away.

Meanwhile, Adam Smith is an evocative reminder that a rational (market-driven) division of labour cannot be accomplished without upheaval. Even in the best-regulated international communities, in the short run one country's industrial gain is likely to be another's loss. In the heady 1960s, when it seemed as if economic growth (and world trade) would continue without interruption, there was endless talk in places such as GATT of the virtues of adjustment assistance — ways of compensating losers for short-term losses. But that talk has now faded. Yet when the danger that nations will barricade themselves behind their trade barriers is greater than ever, the need for such assistance is ironically more urgent than it has ever been.

London and Moscow in hassle

Huxley attacks Soviets on civil rights

A complaint by the Royal Society about the treatment of Soviet scientists by their government has been delivered to the Academy of Sciences of the USSR, and the reply has fallen "far short" of what was promised, and hoped for. This was the most arresting message in the anniversary address to the Royal Society, delivered by Sir Andrew Huxley, the president, last Tuesday. While Sir Andrew reaffirmed his adherence to the society's view that scientists who break their countries' laws are just as liable to punishment on that account as are other citizens, his decision to make public the Royal Society's dissatisfaction with its enquiries of Soviet scientists is a break with precedent and recent practice.

The exchange between the two academies apparently began in May this year, when a Soviet delegation visited the United Kingdom. Sir Andrew explained this week that while not wishing "to sacrifice international cooperation between individual scientists in the hope of persuading the government of another country to change its internal policies", oppressive treatment of scientists by the Soviet government was already "harming scientific interchange" with Britain.

By way of illustration, Sir Andrew said that the funds available each year for exchanges with the Soviet Union are not fully taken up, partly because British scientists are unwilling "to spend time in the USSR because of the oppression in that country of dissident scientists, the refusal of permission to Jewish scientists to emigrate to Israel and recently the revoking of the academic degrees of certain individuals".

The meeting with the Soviet delegation in May this year seems to have been turbulent. Sir Andrew said this week that "we took this opportunity of putting this point forcefully to them". By his account, the Soviet delegation, emphasizing that the Academy of Sciences was not responsible, offered to investigate any cases the Royal Society cared to specify in detail, and to report. The Royal Society then apparently wrote with details of twenty-one Soviet scientists in trouble — seventeen who were being "subjected to various forms of punishment", five refusniks and three who had been prevented from attending scientific meetings in the United Kingdom. The reply from Moscow appears to have arrived only early in November, and to have been unsatisfactory. "Seventeen of the twenty-one scientists named in our letter are not even mentioned, and I take

this as a tacit admission that in most or all of these cases there is no respectable excuse that can be put forward as justification of the treatment they have been given."

Huxley went on to say that he finds it "difficult to accept" the explanation given for the treatment of Aleksander Lerner, who has been refused permission to emigrate to Israel; can "reasons of national security, including knowledge of classified and defence material", apply to one dismissed from his post at the institute of control sciences in Moscow eleven years ago? The Soviet academy also says that V. Kislik is imprisoned for "petty crimes" without saying why he has been deprived of his degrees, and that Y. Orlov is imprisoned for having damaged the Soviet Union politically.

The case of Andrei Sakharov seems, however, to have been the most serious bone of contention. Sir Andrew said that the Royal Society's letter complained that Sakharov was being "persecuted for his work in defence of human rights", and

went on to say that nobody denied that he had criticized restrictions of individual freedom in the Soviet Union, published writing abroad and yet been allowed to remain a member of the Soviet academy.

The Soviet academy's reply seems to have been a deliberate misreading of the complaint. Sakharov's exile to Gork'i is said to be "comparatively mild" in relation to "his actions, which had the aspect of being crimes against the state". Huxley says that the Soviet reply on Sakharov emphasized his dissent from Soviet policies on arms control whereas "most in the West" consider that the "real reason" for his exile is his criticisms of "the illiberal internal regime in the USSR".

Huxley concluded his address by saying that his decision "to express publicly my own feelings" is to be taken as a definition of his position, but that he would not wish to curtail the exchange of scientists between the two countries or restrict visas "to *bona fide* scientists wishing to attend international meetings". ●

French university lecturers protest

Hundreds of French university lecturers marched from the Sorbonne in the centre of Paris to the ministry of national education last Friday, in protest against what they see as a complete abandonment by the minister, M. Alain Savary of any serious — and socialist — reform of the university system.

Much had been — and is — expected of Savary's "law for higher education" to be put before parliament by the end of the year, but the higher education trades unions have not been much impressed so far. Hence the strike.

Backing for the union view, however, may not be particularly strong. A few hundred demonstrators are only a small part of the 44,000-strong community of university teachers, and the union demands for a single kind of lecturer (including professor), with promotion by seniority are seen as unrealistic in many quarters.

Nevertheless the outburst of grievances among the unions has a substantial basis — the enormous blockage of posts within the universities, which according to ministry calculations will see only 158 retirements next year, one third of one per cent of the staff. Thus a common international problem among universities is gravely exaggerated in France, because of a peak in the recruitment of staff in the late 1960s, before and after the "troubles" of 1968.

Moreover, this tinder-box situation has been aggravated by some expectation of relief, after the savage activities of the previous, Giscardian, minister of the universities. But Savary is offering little or no relief, and indeed is piling on the pressure with a reduction in part-time teaching allocations, a lengthening of the

university year, and an insistence that since university staff are civil servants, they must take no more than five weeks' holiday. The number of students would also increase from its present 800,000 to 1 million, said the minister.

Recently the minister further announced to the unions that their dream of a single category of lecturers was unattainable, and that there would instead be a system much like the present, with effectively lecturers, senior lecturers and professors with appointments between grades determined by peer review. That was the last straw, and led the left-wing unions to the strike.

Now, however, the ministry is suggesting that "new propositions" will be made within a couple of weeks — apparently opening all options once more.

This seems to leave the preparation for a new higher education law, one to replace the somewhat rushed law of 1968, in an utter shambles. So many interest groups are at work that it is little wonder that the French education system is sometimes described as the largest political system outside the Red Army.

Moreover, there is a further muddle ahead: the question of appointments and promotions for 1983, previously fixed in late December and early January by the Conseil Supérieur des Corps Universitaires (CSCU), a giant partially-elected parliament of specialist committees. Savary dissolved the previous CSCU, a product of the previous government, a year ago. But this year, without an election policy, each committee has had to be filled by a lottery. It is hoped that the promotions will not be equally random.

Robert Walgate

French science

No big deal

Eighteen months after the arrival of the Mitterrand government in France, with its promise of a great new deal for French scientists, reform of the Centre National de la Recherche Scientifique (CNRS) is complete, or almost complete.

Last week the new statute for CNRS was finally agreed at the Council of Ministers, and published along with the appointments to seven top jobs — some of them new — and the budget details for 1983, showing continued growth of about 13 per cent in real terms (outside salaries). What remains are the elections to the Comité National, the CNRS "parliament" which makes detailed decisions on grant allocations, appointments and promotions in CNRS laboratories. The elections on new (reformed) principles are planned for March.

M. Maurice Godelier, the anthropologist whose appointment as director of social sciences was vehemently opposed little more than a year ago, has now slipped into the role without much protest. His far-ranging report on social sciences to the minister of research and industry (see *Nature* 28 October, p.771) appears to have had a calming effect, and he has been appointed jointly with the previous acting director, M. Armand Frémont, to a combined directorate of social sciences and humanities. Frémont is expected to concern himself with regional issues, while Godelier will be concerned with the higher issues of policy.

Two new directorates have been created: one for information and publications, and another for the application of research ("valorization"), which should strengthen the role of the CNRS in innovation, a principal objective of the government.

The director of valorization will be M.

Jean-Jacques Duby, a 42-year-old mathematician who has had a distinguished career in IBM research and management. M. Duby will consider offering incentives to research groups in the form of additional grants for basic research in return for time spent on applications, a scheme which works well at the Institut Pasteur. CNRS will also seek an increasing role in the research training of third world scientists.

As for the 1983 budget, the net real increase of 7.3 per cent (including salaries), the 320 new research posts (3.4 per cent) and 220 new technicians (1.5 per cent) will be divided unevenly according to the particular needs of each sector. Thus the social sciences receive a large sum for new buildings, to house the home workers; biotechnology sees a start on a new FF20 million (£1.7 million) laboratory for the molecular biology of plants in Strasbourg (which will take three years to complete); and physics sees an allocation for the construction of a new synchrotron radiation source for Orsay, "Super-ACO".

Dividing the budget differently, however, one of the most substantial increases planned for next year comes in medium-sized equipment — instruments costing roughly between FF 1 million and FF 3 million. Here the budget will rise 60 per cent. There is some question, however, whether this sum will survive "regulation" — the mechanism whereby the minister of finance holds back some monies previously promised, subject to good performance of the French economy. This year CNRS did well — losing only FF 50 million in the last analysis, in a budget of FF 6,000 million, but there are fears of worse controls to come — even up to 30 per cent. M. Jean-Pierre Chevènement, the minister for research and industry, is fighting this battle with his ministerial colleagues, in the hope of protecting research from the ravages of the government deficit.

Robert Walgate

Professional liability

Biologists at risk

Washington

As biology comes to have commercial implications, biologists need to insure themselves against lawsuits: their professional lives are getting risky. This seems to be the unwritten message of a brochure now being mailed to members of four US scientific societies concerned with biology and biomedicine.

A "professional liability insurance policy" is being offered for \$62 per year to members of the Federation of American Societies for Experimental Biology, the American Institute for Biological Sciences, the Society for Neuroscience and the Society for Medical Research. The policy is organized by Maginnis and Associates of Chicago and underwritten by the National Union Fire Insurance Company of Pittsburgh.

The policy covers up to \$1 million in costs, in addition to repayment of legal defence costs, for each "occurrence" or incident in which an insured scientist stands accused. It will also pay these costs if the lawsuit arises years after the policy has expired, so long as the events in question took place while the policy was active.

Maginnis officials say the policy covers scientists in situations where they do not normally have insurance cover. An employer such as a university, they note, might be sued for something a faculty member did and then decide the faculty member was negligent and sue him. Or the university might decide that the action in question was outside the faculty member's "scope of employment" and leave him to fight on his own.

The policy would also cover a scientist when consulting outside his regular employment, with a company or individually, whereas his regular employer's insurance would not.

Scientists are covered for advice they give to doctors or nurses about how to treat patients in case the patient later sues the scientist as part of a malpractice suit against his doctors. But the proposed insurance policy does not cover the scientist if he talks to the patient directly, so cancer gene specialists are advised not to hand out free advice to potential patients at cocktail parties.

Arthur C. Gentile, executive director of the American Institute of Biological Sciences, says the policy "covers just about every kind of professional activity". He cites a case where there was an explosion in a laboratory and people who were injured sued. Although the institution that owns the laboratory would probably be liable, the scientist running the laboratory might be named too. "The liability may not be clear and in the meantime you have legal costs" says Gentile.

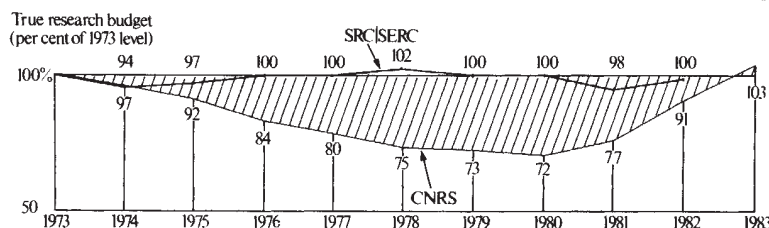
The policy covers someone charged with libel or slander — such as harming another

Budget trends in Britain and France

For Britain, the graph shows the evolution of the budget of the Science and Engineering Research Council, excepting salaries and corrected for inflation by the retail price index. The figures are given as percentages of the 1973 budget. This indicator of "true money for science" is almost exactly constant over the decade. By contrast, the roughly equivalent Centre National de la Recherche Scientifique in France shows a strong decline in the same

quantity over the 1970s, followed by a small rise in 1981 (the last year of the previous government). The strong increase in the past two years indicates the impact of the Mitterrand era. But the 1983 budget will only just pass that of a decade earlier. The continuation of the slope to 1985 is broadly guaranteed by the "law for research" of July this year but may be affected by the uncertain state of the teetering French economy.

Robert Walgate



scientist's reputation by overheated statements in the classroom or in published articles.

In the court fight between Hoffman LaRoche and the University of California at Los Angeles (UCLA) over how Hoffman LaRoche obtained a KG1 cell line belonging to the university, the company also named individually two UCLA scientists, David W. Golde and H. Philip Koeffler. It also named Golde separately and accused him of slandering the company. The university is conducting the entire case, including the parts in which the individuals are named, and carrying the costs.

Golde's view is that he therefore does not need the insurance policy now being offered. (Moreover, the university has a malpractice policy that covers him because he is also a physician.) But he warns that "there is not always a 100 per cent

congruence of interests between a faculty member and his university in this kind of situation".

As for the likelihood that a university, after defending one of its faculty, would turn round and sue him, Harvard's general counsel Daniel Steiner says it is "far-fetched".

The brochure does not mention the problem of liability for new organisms created through gene-splicing: would the policy cover, for example, a scientist who invents a bug that eats Manhattan, as one observer phrased it? But the policy would in any case be of only marginal use in that situation. For one thing, the likely damages would exceed the \$1 million liability limit. Moreover it does not cover any "punitive" damages — the extra penalties a judge can impose to deter the scientist from, say, turning his bug loose on Chicago.

Deborah Shapley

of the hardy Novosibirskaya-607 winter wheat and a strain of sheep with a 50 per cent increased wool yield to the hormonal control of plant growth and pig fertility and the protection of grain by irradiation.

Such a media dialogue is a well-known ploy in Soviet journalism; a new drive is launched and backsliders rebuked, and then a selected representative of those under attack states his case. The matter may then be dropped (on the assumption that those who deserved criticism will have profited by the warning), or else may go on to become a full-scale press campaign of criticism, culminating in a new party and government decree.

What was not routine in this case was, of course, the death of Mr Brezhnev, which necessitated the convening of the Central Committee plenum, and the latter's announcements of, effectively, new party directives. For Dr Aleksandrov, who, only a few days before, had on behalf of Soviet science, made one of the five valedictory speeches at Brezhnev's funeral, it was time to make his position clear. The scientists had produced the designs, he said, and in some cases they were implemented "quite rapidly and effectively". In many ministries, however, the need for updating technology had been discussed for years, but scientists had not yet been able to establish "proper working contacts". The result was waste, inefficiency and delays. The firm reiteration of the problem by the plenum, he hoped, would get things moving again. He stressed, however, that if the scientists' work is not to be wasted, economic managers should be more accountable for their work in implementing new technology.

Vera Rich

Soviet technology

Red tape blamed for delay

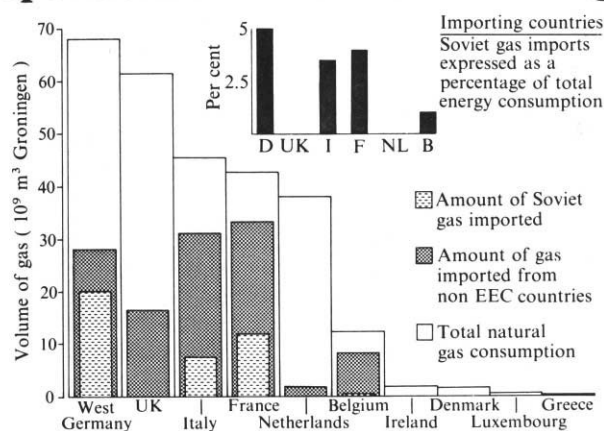
Soviet bureaucrats — not scientists — should be blamed for persistent delays in the introduction of new technology, Academician Anatolii P. Aleksandrov, president of the Soviet Academy of Sciences, said last week. He was commenting on the recent plenum of the Central Committee of the Communist Party of the Soviet Union, which, in his words, had put "new pressure" on the drive for the more rapid implementation of new technology. This drive was a major issue in the policy of the late President Leonid Brezhnev, and the stress that the plenum put on it seems an intensification rather than a change in policy. Dr Aleksandrov, however, may find his own position under the new administration more vulnerable. It is widely believed that, although he was elected by the academy in a secret ballot in the regular way, his candidacy was known to enjoy the support of Mr Brezhnev, and there has been considerable speculation that with Mr Brezhnev gone, Dr Aleksandrov may now be replaced as president of the academy.

In any case, as president of the academy, Dr Aleksandrov holds considerable responsibility for the technological modernization programme, since, during the last five-year plan, the academy was given a major coordinating role and charged with establishing and strengthening links between researchers and industrial customers. This move followed a wave of criticism of scientists for not pulling their weight with those science centres which had been specially created to serve the needs of local industry singled out for special blame — starting with the relatively small Far Eastern Science Centre, and working up to the prestigious Siberian branch of the Academy of Sciences. The scientists retaliated vocally, blaming the administrators and managers who failed to

implement their recommendations.

Scientists were again blamed earlier this autumn. On 20 October, *Pravda* carried an article by two senior officials alleging that research potential was being wasted and misapplied in the various institutes of the Academy of Agricultural Sciences. Three days later, in an interview on Moscow radio, Dr Vladimir Shumnyy, deputy director of the Cytology and Genetics Institute of the Siberian branch of the Academy of Sciences, tacitly denied the accusation, citing an impressive list of joint achievements of the Siberian branches of the two academies, from the development

EEC dependence on Soviet national gas



The figure shows the dependence of the European Economic Community (EEC) on imports of natural gas with particular reference to the Soviet Union. Natural gas now accounts for 18 per cent of the total energy consumption of the EEC, compared with just 7 per cent in 1970. But fears that demands for imported gas may lead to an increased dependence on Soviet supplies appear to be unfounded, as shown by the situation envisaged for 1990 which is illustrated here. Gas imported from non-EEC countries will account for 44 per cent of the total amount of gas required. Of this, 19 per cent will come from the Soviet Union, accounting for only 4 per cent of the total energy consumption. Other non-EEC suppliers will be Norway (12 per cent of natural gas imports), Algeria (10 per cent) and Libya (1 per cent), and 2 per cent will come from unspecified sources.

Plant breeding

Genetic diversity threatened

Washington

The International Board for Plant Genetics Resources (IBPGR) is in a race against time to preserve the diversity of germplasm needed for future plant breeding programmes, but its success is threatened by cutbacks in support for international agricultural research.

"If the work is not done in the next 5–10 years, we're finished", Dr J. Trevor Williams, executive secretary of IBPGR, said here last week. The spread of a few high-yielding varieties of many crops "forces the time-scale", and these new varieties are quickly replacing the diverse traditional varieties which have evolved over thousands of years and which represent the essential raw material for the development of future strains. Wild varieties, too, are threatened by such changes as forest clearing in the tropics.

IBPGR, which is part of the Consultative Group on International Agricultural Research (CGIAR), has since 1974 been sponsoring expeditions to gather samples of wild and traditional varieties of key crops and has coordinated efforts to provide storage facilities to preserve these germplasms. But, as a recent report from the CGIAR secretariat notes, "germplasm activity can be discontinued at short notice and, over the short term, its lack is not immediately missed. Over the longer term, its lack will be serious."

CGIAR's support for 1982 is expected to fall about \$14 million short of even its low-budget option of \$166 million for the year; and given the vulnerability of germplasm activities, IBPGR is "questioning whether

it can continue to fund collecting missions or equipment for germplasm storage", according to the report.

IBPGR has sponsored some 250 missions, which have collected about 100,000 new seed varieties. The board has also helped to coordinate the various national germplasm programmes, and has been particularly active in the generally overlooked area of tropical fruits and vegetables. According to Dr William Brown of Pioneer Hi-Bred, a seed company in Des Moines (Brown is also chairman of two crop advisory committees to IBPGR), the board has also played an important part in providing "insurance" in the germplasm resource system. "A real effort has been made to have them stored in more than one place."

The importance of the germplasm collection and storage activities has been underlined by several recent developments. A new variety of rice developed by CGIAR's International Rice Research Institute in the Philippines, for example, included a wild variety of rice (*Oryza nivara*) in its breeding to confer resistance to a disease known as grassy stunt. This new variety, called IR36, has been adopted with remarkable speed, displacing traditional varieties. Almost 11 million hectares have been planted with IR36, which now accounts for more than 60 per cent of the rice grown in the Philippines, Indonesia and southern Vietnam. And, as Dr Williams pointed out, once the traditional varieties are gone, "there is no way you can get them back".

Stephen Budiansky

Biotechnology

More help

More help for British biotechnology is on the way. Last week, Mr Patrick Jenkin, Secretary of State for Industry, announced that his department will spend £16 million in the next three years to promote innovation and awareness in biotechnology among British companies. The scheme, modelled on an earlier scheme that promoted microprocessor technology, should go some way to answering persistent criticism of the British government for failing to provide effective support in this area.

In July, a House of Commons select committee complained that the government had failed to protect university research in biotechnology against the reduction of support from the University Grants Committee (UGC) in 1981. Although UGC had set aside £1.2 million for selected universities to support research in biotechnology, the select committee feared that much of this would be used to make good deficiencies incurred in earlier cuts. (The final allocation of this sum has yet to be announced.)

The new scheme from the Department of Industry is intended to complement UGC and research council support, and will be coordinated by a new biotechnology unit set up at the Laboratory of the Government Chemist. A graded system of grants will help companies to commission strategic studies, feasibility studies and specific problem-solving studies by independent consultants. There will also be cash grants for innovative demonstration plants as well as for research projects funded jointly with industry at selected centres. The department also promises a "general strengthening of the biotechnology infrastructure", which will include support for an information service and culture collections.

Mr Jenkin explained that the government was introducing the measures because potentially commercial biotechnology investments entailed high risk and long lead-times. Another reason may be growing concern that British scientists, frustrated by lack of opportunities for research, are being lured abroad. Other recent publicly supported initiatives include a biotechnology research centre at the University of Leicester, supported jointly by the Science and Engineering Research Council and several industrial companies.

Mr Jenkin maintained that British support for biotechnology now compares favourably with that in other countries, an assessment that has been challenged by opposition Labour MPs: one described the scheme as "no more than a drop in the ocean". The Confederation of British Industry has cautiously welcomed the new plan, although a spokesman said of the £16 million figure, "obviously we would like to see it higher".

Tim Beardsley

"Creationism" thwarted

Washington

A Louisiana law requiring the teaching of "creation-science" in public schools was struck down last week in the federal district court. Judge Adrian Duplantier ruled that the law, virtually identical to the Arkansas statute that was overturned last year, infringed upon the authority of the state's Board of Elementary and Secondary Education to set school policy.

In contrast to the Arkansas case, the judge made no finding of fact on the validity of "creationism" as science. "We didn't even deal with the facts", said Burt Neuborne, legal director of the American Civil Liberties Union (ACLU), which brought the suit. "We just said the wrong government agency acted."

Under the Louisiana constitution, the education board has authority to "supervise and control" the public schools; the legislature is charged to "establish and maintain" the schools. The judge ruled that in specifying a

detailed course of study, the legislature overstepped its authority. The board, which was originally named as a defendant in the suit, joined with ACLU in arguing that its prerogatives were being usurped by the law. The board, composed of professional educators, is unlikely to institute a "creation-science" teaching requirement on its own. Louisiana's attorney general, however, immediately announced that the state would appeal the decision to the federal circuit court of appeals.

Neuborne suggested that similar separation-of-powers arguments could be raised under the constitutions of other states that pass "creation-science" laws. Georgia and Alabama may do so next year. "Separation of powers is an incredible mine of judicial resources", Neuborne said — and may also be an inexpensive one to tap. "The Arkansas case cost us \$1.2 million", he said. "This one cost us eight sheets of yellow legal paper."

Stephen Budiansky

Drugs for developing countries

Uncharitable stance from Oxfam

In the poorer countries of the world, millions of people suffer unnecessary ill health or die because simple drugs such as antibiotics are not available. But to what extent have the marketing methods of the multinational pharmaceutical manufacturers contributed towards the misuse and waste of drugs in the Third World?

According to the British charity Oxfam, the manufacturers have blatantly promoted inappropriate and harmful drugs in developing countries, which lack the tight governmental control on drug sale and distribution now normal in wealthy nations. And last week Oxfam renewed its attack on the industry by publishing a new study, "Bitter Pills" by Dianna Melrose. Despite a rider from the charity's director-general Brian Walker, recognizing the "valuable contribution of the research-based drug industry to world health", the study casts the industry as the villain of the piece.

But the industry, in the shape of the Association of the British Pharmaceutical Industry (ABPI), says that until it proves possible to build up adequate health-care systems in the poorer countries so that appropriate drugs can be properly prescribed and distributed, there is little the industry can do.

Abuses cited by Oxfam fieldworkers include: the maintenance of "high prices" for pharmaceuticals; inaccurate or misleading promotional literature; and aggressive promotion of "non-essential" drugs at the expense of more vital products. A common complaint is that over-prescription is encouraged — so, for instance, an impoverished patient with a bacterial infection might be prescribed a tonic and a multivitamin preparation in addition to the vital antibiotic. Since the antibiotic is likely to be the most expensive of the three, and the patient cannot afford them all, the cheaper, heavily promoted tonic and multivitamins may be preferred.

One of Oxfam's answers to over-prescription is to limit the number of drugs on sale in developing countries. The World Health Organization (WHO) has issued a list of approximately 200 "essential" drugs (many others being deemed "useful") for developing countries, and in June Bangladesh's military government took the

dramatic step of banning the sale of more than 1,700 drugs termed "harmful" or "unnecessary" and allowing only 150 "essential" drugs from the WHO list to remain on the market (*Nature* 15 July, p.219). Local and international drug manufacturers objected, but although there have been some additions to the list of permitted drugs after strong protests, the government is to go ahead with its plan.

Thus events in Bangladesh over the next few years should show just what can be



achieved by severely restricting drug supply in a country where many of the problems of distribution, prescribing and poverty persist.

There seems little doubt that a policy based on greater promotion of WHO's 200 "essential" drugs for the developing world will be of some benefit. About 50 drug companies are already cooperating with WHO by promising favourable prices for listed drugs, although further moves like that in Bangladesh actually to prohibit the sale of products not listed would not be likely to encourage further cooperation.

Oxfam, though, seems likely to remain unconvinced. WHO is seen as well intentioned, but "a toothless tiger", in the words of Oxfam's Dianna Melrose, and there is mutual distrust between Oxfam and the drugs industry. Were the two camps to stress the many views they hold in common rather than relish the differences, the world's developing countries could only benefit.

Charles Wenz

British universities

Agony at Aston

The legal quirk that provides security of tenure for full-time academics at British universities is likely to be tested in a forthcoming battle between the University of Aston in Birmingham and the Association of University Teachers (AUT). The outcome will set an important precedent for other universities.

Aston was told in July 1981 by the University Grants Committee (UGC) that its budget would be reduced by 30 per cent in the succeeding three years. A large-scale programme of voluntary redundancies has already been approved, but now the university's council has voted by a narrow margin (19 votes to 17, with one absentee and one abstention) to effect compulsory redundancies under the 1975 Employment Protection Act. It is thought that 41 academics (chiefly in the fields of engineering and mathematics) could be given notice before Christmas.

The union, according to its general secretary, Mr Laurie Sapper, is prepared to fight all the way. An application to the High Court for a writ to prevent the redundancies is being considered. The grounds of the action are that compulsory redundancies would be contrary to the charter and statutes of the university which (like those of most universities in England and Wales) specify that "no member of staff . . . shall be removed from office . . ." except on grounds such as physical or mental incapacity and immoral, scandalous or criminal behaviour; there is no mention of financial circumstances as a cause for dismissal.

AUT has sought access to the university's accounting figures so that it can carry out its own examination of the university's finances, but Mr Sapper says that Professor Frederick Crawford, Aston's vice-chancellor, has failed to respond to AUT's letters. AUT has accordingly applied for access under the Employment Protection Act.

Aston is primarily technology-based, and has forged extensive links with local industry. The university says that the unemployment rate among its graduates is one of the lowest in Britain.

UGC recommended in July 1981 that student numbers in several science subjects should be reduced and that courses in biological science and architecture should be discontinued. Aston has already axed its courses in human communication and behavioural sciences. The number of science students has fallen from 3,100 in 1980-81 to 2,732 now, while arts student numbers have fallen from 1,383 to 1,340, but further reductions have still to be made. Full-time academic staff, of whom there were 522 in August 1981, will have to be reduced to 357 by July 1984.

At the same time, the university plans to recruit extra staff in some subjects, notably



Distribution problems — penicillin exposed to heat and sunlight for sale in a market in Upper Volta.

for its Management Centre, which received special praise from UGC. It is also determined to maintain a biology presence, possibly by means of a new biotechnology unit. But staff of the existing biology department have refused to examine that plan on the grounds that it would probably involve abandoning the university's established expertise in applied biology and also lead to redundancies.

Dr Richard Etheridge, a senior lecturer in production technology at Aston and a member of the AUT national executive, says that the university staff feel that the vice-chancellor is imposing redundancies with unseemly haste. The Department of Employment was notified of the proposals (as required by employment law) on 18

November, and redundancy notices could be issued as early as 22 December. Several attempts by academic staff at the most recent meeting of the university senate to introduce motions urging reconsideration or delay were blocked by the vice-chancellor on the grounds that these were not procedural matters.

The issue is certain to be raised again at

the next meeting of the university council on 14 December. AUT is planning to lobby council members on that day, urging them to reconsider, and has threatened to tell its members elsewhere to boycott the university, depriving it of external examiners and putting existing procedures for the award of degrees in hazard.

Tim Beardsley

Spina bifida

Controversy over tests continues

Heart-searching and a little politicking about the UK Medical Research Council's proposed controlled trial of the utility of folic acid and other vitamin supplements as a prophylactic of spina bifida continues, fanned by the appearance of a blank space in last week's edition (see *Nature* 25 November, pp.302 and 310). What follows is an account of the origin and the course of this controversy.

The immediate stimulus of present interest in a controlled study of folic acid supplementation appears to have been the publication in February 1980 (*Lancet* i, 339) by Professor R.W. Smithells of the University of Leeds and seven collaborators of an article entitled "Possible prevention of neural-tube defects by pre-conceptional vitamin supplementation". That reported on a study in which women who had previously given birth to a spina bifida child were offered a vitamin preparation containing folic acid. This report appears to have prompted the UK Department of Health to enquire, in 1980, of the Medical Research Council whether a controlled study of the efficacy of folic acid would be feasible and desirable.

Spina bifida results from the incomplete closure of the neural tube in the early stages of embryonic development, and is thought to date from 15 to 20 days after conception. The other principal neural tube defect in human beings is anencephalus. Animal models for the study of neural tube defects, among which the curly tailed mouse stands out, have not so far provided a clear picture of the mechanism by which the neural tube is closed in normal development, although they do clearly show an interaction between genetic and environmental factors in the causation of neural tube defects (see Morriss, G. *Nature* 284, 121; 1980).

That a dietary deficiency of some kind may be involved in the occurrence of neural tube defects in human births is suggested by the correlation between the incidence of the defects and the social class of the mothers concerned. (The incidence is greater among less prosperous families.) Another pointer of the same kind is the increased incidence of defective births among children conceived in the late winter and early spring.

The geographical variation of the incidence of these defects is, however, harder to account for. In the British Isles, the incidence of both defects is greatest in the west (Ulster and the Republic of Ireland),

amounting to 9 per 1,000 in Belfast, 7 per 1,000 in South Wales, 4 per 1,000 in Birmingham (central England) and 2 per 1,000 in south-east England. By contrast, in France the incidence of the defects is less than 1 per 1,000, while the defects are almost unknown among some of the poorest populations of the world. In North America, the incidence of the defects appears to decrease from a maximum on the east coast of about 2 per 1,000 to less than half as much on the west coast. There appears to be a puzzling cyclical variation of incidence in the United Kingdom with a periodicity of twenty years or so.

That genetic factors influence the occurrence of neural tube defects is borne out, for example, by studies in South Wales in the 1960s. In one survey (Carter, C.O. *et al. J. med Genet.* 5, 81; 1968), 5.2 per cent of the siblings of an affected mother were found to produce neural tube defective children, compared with the general incidence of 0.77 per cent in the population as a whole. The risk of neural tube malformations is, puzzlingly, greater among first births and twice as great among females as among male children. One epidemiologist said this week that there can be no other congenital malformation whose incidence provide so many clues but whose causation is still in doubt.

The notion that a deficiency of folic acid may be partly responsible appears to stem from measurements by Smithells and his colleagues suggesting that red-cell folic acid and leukocyte ascorbic acid (vitamin C) might be deficient in mothers giving birth to defective children. These measurements led to Smithells's clinical trial of the efficacy of vitamin supplements including folic acid in the prevention of recurrences of neural-tube defective births.

Women who had previously given birth to spina bifida children were recruited to the study principally from genetic counselling clinics, but also by referral from general practitioners, and were offered a course of treatment with a multivitamin preparation containing folic acid. Among those who accepted, 137 gave birth within the period of the study, as did 187 of those who were found already to have become pregnant again (or who declined). In the first report of the study, among 140 children born to the mothers supplemented with folic acid, there was only one neural tube defect. But there were

Object lesson

Brussels

In an unusual demonstration of its powers, the European Commission has questioned West Germany's second energy research programme, which is suspected of contravening EEC's rules on competition. The West German government has six weeks in which to reply.

The programme in question, which covers the period up to 1985 and costs DM13,465 million (£3,370 million), was in fact approved last year and is already operating. Nevertheless, the Commission complains that it was not informed in due time about either this programme or the last one and points out that the programme is a form of state aid and so is subject to the Commission's approval.

In its letter, the Commission declares that it wants to examine in greater detail the extent to which the programme fits in with the Community's energy policy aims. The fact that much of the money will be set aside for large-scale projects, particularly in fields such as fast-breeder technology, largely carried out in Germany by private companies, leads the Commission to question whether or not the aid is leading to a distortion of competition within the Community. And second, the Commission is concerned about whether only German companies and not, say, the foreign subsidiaries of companies from other parts of EEC, will be able to take advantage of the loans and subsidies.

This is the first time that the Commission has raised objections to a national research programme and its action is being interpreted in Brussels as a sign that research programmes should not be used by national governments as a discreet form of industrial featherbedding. If the Commission is not satisfied with the reply from Bonn, and is successful in pursuing its action, a similar letter might soon be on its way to Paris.

Jasper Becker

12 neural tube defects among the 192 infants born to the mothers in the untreated group.

More recently, Smithells and his colleagues have extended their study to include more than 1,000 women with a history of neural-tube defective births, self-selected into two roughly equal groups. They report that the incidence of a second neural-tube defective birth was 0.6 per cent (based on three births) in the group of supplemented mothers, but 4.7 per cent among the children born to untreated mothers.

This seven-fold difference apart, the bearing of the Smithells study on the question of whether vitamins including folic acid were an effective prophylactic of the recurrence of neural tube defects is disputed. First, and most seriously, because the study group is self-selected, the results may be seriously biased, and in the direction reported. For if mothers of higher social class are less likely to decline a course of vitamin supplementation, and are inherently less at risk of producing neural-tube defective children, the results of the trial may partially be explained.

On this point, Smithells argues that the data now accumulated include 44 women with more than one neural-tube defective child who have agreed to accept vitamin supplements and 52 women with two or more defective children who had not joined the trial. (The first group produced one defective child, the second five.) To Smithells, however, the significance of this development is to give the lie to the criticism of his earlier work that mothers accepting vitamin supplementation select themselves in consonance with their tendency to produce neural-tube defective children.

Another difficulty is that the catchment area from which the participants in this study have been drawn includes Ulster, where the natural incidence is large, and the south-east of England, where it is small. Recruitment to the programme of vitamin supplementation seems indeed to have been more successful in the south-east of England (50 per cent) than in Ulster (25 per cent).

It seems also to be agreed that the Smithells trial does not point unambiguously to folic acid rather than some other components of the vitamin preparation used as the putatively effective prophylactic. The preparation used (obtainable in the United Kingdom only on prescription) is manufactured by Beecham Pharmaceuticals Limited, is known as "Pregnavite Forte F" and contains a mixture of vitamins (including ascorbic acid) as well as iron and calcium salts.

During the past year, the Medico-Social Research Board in the Republic of Ireland has mounted a trial initially intended to throw light on this point. With the collaboration of several Dublin maternity hospitals, the board is sponsoring a trial in which three groups of mothers with a

history of neural-tube defective children are being offered Pregnavite Forte F, Pregnavite Forte (without folic acid) and folic acid alone. Dr Geoffrey Dean, the board's director, said on the telephone last week that "everybody is convinced that the Smithells theory is not proven". The Dublin study also suffers from the defect that there is no objectively chosen control group with which the effects of vitamin supplementation can be compared.

In Dublin there is some resentment that in the past few months Beechams (which had been supplying the vitamin preparations free of charge) has written to say that it cannot continue this provision (see box). Dr Dean hopes that, nevertheless, it should be possible to recruit enough mothers with a history of neural-tube defective children for a significant result to be available within two years.

In Britain, the interest of the Department of Health in 1980 prompted a seminar in December of that year at which several interested and disinterested parties discussed the conduct of a clinical trial intended to yield significant and decisive results, in which the participants would again be mothers with a history of neural-tube defective births.

Beecham's supplement

The decision of Beecham Pharmaceuticals Ltd to go back on its earlier agreement to supply "Pregnavite Forte F" for the Irish trial free of charge, which has caused some ill-feeling in Dublin, is said to have been prompted by the company's internal ethical committee. According to Mr J. B. Diamond, chairman of Beecham (UK) Ltd, the ethical committee includes "secretaries and people like that".

Mr Diamond explained that his company had supplied material for the trials carried out by Professor R.W. Smithells and his associates over a period of some years, and also for the Irish trials. But "when it became clear that the [proposed] trials raised great ethical problems", the committee had decided that it would be wrong for Beecham to supply a placebo for the MRC trial and, by extension, to supply a vitamin supplement containing folic acid for other trials.

On the question of the safety of folic acid as a dietary supplement, Mr Diamond said on the telephone this week that the company had been marketing Pregnavite Forte F since 1952, before the licensing of drugs became the statutory responsibility of the Department of Health on the advice of the Committee on the Safety of Medicines. As such, the product has a licence "as of right", which may nevertheless be called in question whenever the committee considers that there has been a significant change of usage.

The most controversial feature of this proposal is that there should be a true control group — a group of women selected double-blind to whom only a placebo should be administered. Three other groups would receive respectively a multivitamin preparation, a multivitamin preparation with folic acid, and folic acid without other vitamins (but with minerals). The folic acid dose would be 4 mg a day, roughly ten times the amount of folic acid in the Beechams preparation.

The Medical Research Council, which apparently has indications that 20 per cent of the 3,000 women it would need in a full trial are willing to take part, has run into the following snags.

- Several Members of Parliament have been asking questions of the Department of Health about the conduct of the trials. They include Mr Roger Thomas, Mrs Gwyneth Dunwoodie and Mr Leo Abse. The common thread in their questioning has been to ask why their constituents should be denied an apparently effective treatment.

- On some legal views, informed consent by participants in the trial would not necessarily absolve the physicians concerned from legal action of the birth of damaged children.

- The amounts of folic acid planned for the MRC trials (ten times the amounts included in Pregnavite Forte F) have suggested to some that the procedure could be dangerous. While there is no evidence one way or the other, prudence dictates caution.

- If trials should show that folic acid, alone or in conjunction with other vitamins, should be effective at preventing recurrences of neural-tube defective births, it is by no means clear how that knowledge would be used for the prevention of defective first births.

In the circumstances, the opinion about the proposed trials of scientists working in the field understandably ranges over the whole spectrum. Some hold that the Smithells case, if not fully proven, is sufficiently convincing to make the use of a placebo in the trials unethical. Others would avoid the uncertainties occasioned by the choice of a dose of 4 mg a day by arguing that the amounts used should be smaller until animal tests have been carried out.

On the other hand, there are many who hold that the Smithells investigations raise so many more questions than they answer that a trial with placebo is probably necessary. The ideal would obviously be to ring the changes on all the vitamins administered, but this is probably unpracticable in the short term, which is why some argue for detailed physiological investigations of smaller numbers of women at risk.

Professor R.W. Smithells himself hopes that the Medical Research Council "will decide to carry out a trial" at its next meeting on 7 December. ●

CORRESPONDENCE

Human cancer genes

SIR — In one respect, your News and Views article was startling (*Nature* 11 November, p. 103). You have turned upside down the relative values I attach to the different branches of cancer research. You attribute to me the belief that "the prime objective should be to understand the phenomenon of malignancy . . . (rather than to) concentrate on identifying the causes of cancer". This is almost the exact opposite of what I think. In the earlier paper of mine that you cite as evidence of my beliefs (*Nature* 289, 353–357; 1981), I said that although we know the causes of cancer of the skin and lung, "the preventable causes of the other common cancers are less clear-cut and it may be difficult to identify them until much more is known about mechanism". I go on to say that "eventually the techniques of nucleic acid chemistry should allow us to itemize all the differences in nucleotide sequence and gene expression that distinguish a cancer cell from its normal counterpart, and perhaps at that point the steps involved in carcinogenesis will cease to be in doubt". That was written in 1980, and it still seems a reasonable forecast. Indeed, Logan and I finished the News and Views article that immediately follows your piece with the sentence "But we can at least feel that the technology for studying genes and their patterns of expression are now advancing so fast that the molecular biologist will soon be telling the epidemiologist what to look for".

At the risk of being tediously repetitive, let me say it once again. I believe that the only promising strategy, in the foreseeable future, is to learn what are the causes of cancer and then learn how to avoid them. For, in the absence of prevention, all we are left with is treatment — and that seems, at present, to be an intractable problem.

JOHN CAIRNS

Harvard School of Public Health,
Boston, Massachusetts, USA

Cohen-Boyer patent

SIR — In reading the article about Stanford's response on the "second Cohen-Boyer patent" to the US patent office (*Nature* 11 November, p. 95), one is struck by the editorializing by the author. The matters brought up by the patent examiner were answered very thoroughly by the patent attorney, and interested readers may obtain the response from the patent office to make their own judgements.

We do, however, wish to respond to the issue raised concerning the possible co-inventorship of Dr Robert Helling, as the article makes it appear the universities have deliberately sought to exclude Dr Helling from being a co-inventor. The facts are these.

Dr Helling has refused to provide evidence supporting his assertion in the press that he is a co-inventor or otherwise communicate with the patent attorney, Mr Rowland, whose professional responsibility it is to determine that the true inventors are indicated on any patent application he files. Under the rules of the patent office, Mr Rowland now is unable to change his original determination of inventorship unless Dr Helling provides the requested information. Issuance of the patent

is not affected by this matter of inventorship. If Dr Helling is an inventor, the universities want his name on the patent. Patent office procedures allow for filing of a reissue of a patent to change named inventors.

Lastly, we are disappointed that documents submitted to the patent office regarding inventorship and available to the author of the article were ignored. These would seem to contradict directly statements to *Nature* about Dr Helling's role in the invention.

NIELS J. REIMERS
(Director)

Technology Licensing,
Stanford University,
Stanford, California, USA

Stephen Budiansky responds: The documents Mr Reimers refers to are a short summary by Dr Helling of his research activities, filed with the University of California when he left there in 1973, and submitted to the patent office last month by Mr Rowland. I omitted mention of it because it does not bear on the issue of inventorship. In it Helling states that he "participated in developing a procedure for coupling DNA fragments from any source to bacterial plasmids"; this neither supports nor contradicts his claim that he made a significant contribution to the original conception.

Mr Reimer's assertion that Dr Helling refused to provide information to Stanford's attorney is, I understand, a matter of dispute between the parties, and is in any case academic now that the issue of inventorship is under discussion between Dr Helling, Stanford and the patent office.

Green ecology

SIR — I doubt that I shall be able to dispel the impression gained from the disparaging article "No future for the Greens" (*Nature* 28 October, p. 766) but I would like to correct some factual inaccuracies.

The Ecology party conference at Bridlington was not its first public meeting; it has held annual conferences since 1974 and additional policy conferences in 1981 and 1982. All these conferences have been open to outside observers.

Your reporter has also made an error common to ill-informed commentators in suggesting that the Ecology party is solely concerned with environmental issues. As described in its manifesto the Ecology party has policies on all those aspects of life that concern the other British political parties. The difference lies in the application of ecological principles to these concerns in much the same way as the Labour party, for example, would claim to apply socialist principles.

It is correct to suggest that the German Green party (*Die Grünen*) emphasized environmental issues in the recent elections but this does not imply that such issues are the sole concern of the Ecology party. That the Ecology party does not only consider environmental issues may partially explain the greater impact of *Die Grünen*, but the difference between British and German electoral systems is probably more important. The third factor hindering recognition of the Ecology party is the inaccurate reporting in the British media of which the abovementioned article is an example.

Southport, UK

R. W. SMALL

Science now and then

SIR — In the course of reviewing Sir Peter Medawar's "Pluto's Republic" (ref. 1), Stuart Sutherland commented favourably on the work of K. Popper but less favourably on the work of P. Feyerabend.

Despite Feyerabend's currently anti-Popperian stance Popper² has recently drawn a distinction between yesterday's "great science" and today's "big science", indicating that the latter should be attacked along Feyerabendian lines. Popper³ concluded: "But new ideas should be regarded as precious, and should be carefully nursed; especially if they seem to be a bit wild."

For some of us who were disenchanted^{4,5} with the "big science" of the previous decade, Feyerabend was the philosopher of science. However, there is evidence that "big chemistry" (at least) is becoming more rational and more tolerant of low status authors^{6,7} who publish critical reviews. Thus, Popper's position, strengthened by the recent publication⁸ of "Postscript", is becoming increasingly more attractive.

It is currently considered^{9–11} that the "chemical revolution" initiated by Lavoisier originated from a Feyerabendian assault rather than from a Popperian "problem situation" or a Kuhnian "crisis". Similar judgements^{12,13} are also accorded on Liebig's "biochemical revolution". In both cases the innovators relied on a mixture of empirical evidence and unsupported speculation to justify their theories.

Fruton¹⁴ concludes: "Although most of Liebig's physiological speculations were rejected by several contemporary critics and soon became untenable, there can be little doubt that they stimulated the study of physiological processes from a chemical point of view. The public position that he had won through his achievements in organic chemistry, both as an investigator and as a teacher, assured a respectful hearing for his bold hypothesis. In seeking to prove or disprove Liebig's ideas, his defenders and critics were spurred to examine more closely substances (such as the proteins) present in living organisms . . ."

F. M. AKEROYD

Bradford & Ilkley Community College,
Bradford, UK

1. Sutherland, S. *Nature* 300, 118–119 (1982).
2. Popper, K. R. in *Scientific Revolutions* (ed. Hacking, I.) 89, 96, 106 (Oxford University Press, 1981).
3. Popper, K. R. in *Scientific Revolutions* (ed. Hacking, I.) 97 (Oxford University Press, 1981).
4. Ninio, J. *Trends biochem. Sci.* July, vii (1981).
5. Akeroyd, F. M. *Trends biochem. Sci.* 7, 241 (1982).
6. Ginsburg, D. *Acc. chem. Res.* 15, 199 (1982).
7. Akeroyd, F. M. *Chem. Br.* 18, 481–482 (1982).
8. Popper, K. R. *Postscript to the Logic of Scientific Discovery* (ed. Bradley, W. W. 111) (Hutchinson, London, 1982).
9. Ihde, A. J. in *Joseph Priestley* (eds Kieft, L. & Willeford, B. J.) 77–78 (Bucknell University Press, 1980).
10. Teich, M. *Ambix* 29, 17–21 (1982).
11. Fisher, N. *Br. J. Hist. Sci.* 14, 214–216 (1981).
12. Holmes, F. L. Introduction to the facsimile edition of Liebig's *Animal Chemistry* (Johnson Reprint Corporation, New York, 1964).
13. Fruton, J. S. *Molecules and Life*, 267–269 (Wiley, New York, 1972).
14. Fruton, J. S. *Molecules and Life*, 272 (Wiley, New York, 1972).

NEWS AND VIEWS

New lease of life for pulsars?

The discovery of the fastest pulsating radio star so far has surprised observers, and may yet catch out the theorists.

Pulsating stars are becoming fashionable again. That at least is the import of a report in the *New York Times* on Saturday last week, recording the discovery that a well-known but always puzzling radio source, known as 4C21.53, contains as part of its structure a rotating object of some kind spinning at a remarkable 642 times a second. In the past few weeks, most radiotelescopes have been pointed towards the position of the object at Right Ascension 19h 37m 28.72s, Declination 21d 28m 01.3s.

The discovery, made with the help of the radiotelescope strung with wires across the extinct volcanic caldera at Arecibo in Puerto Rico, is credited by the *New York Times* to Dr Carl Heiles of the University of California at Berkeley. The discovery is obviously a landmark in astrophysics because the pulsations of the radio signals are at least twenty times more rapid than those of the most rapidly pulsating star known so far — the supernova remnant more or less at the centre of the Crab nebula. The *New York Times* rightly says that such an object would naturally be taken for a newly formed pulsating star, in which case it might be thought that the radio signals would be accompanied by optical evidence of an expanding shell of matter of the kind supposedly flung off in a supernova explosion and thus travelling through interstellar space at 10,000 km a second or thereabouts.

On the conventional view of the formation of neutron stars as the condensed remnants of roughly one solar mass left over in the explosion of a star perhaps seven or eight times as massive, the rate of rotation of the new pulsating star should also rapidly decay through the conversion of rotational energy into magnetic flux and radiation. The *New York Times* reports that the failure so far to measure such a deceleration is a sign that the pulsating star may not be a now conventional rotating neutron star, but it is probably too soon to jump to that conclusion. The observation of such an object is not child's play, while this particular one has been recognized as a pulsating radio source only since the first half of November (when astronomical observers were informed through the international telegram service) although, with hindsight, it is clear that evidence of pulsation was first obtained during observations in September. The first full account of this discovery, together with a report on what is visible in that patch of sky and a first attempt at an interpretation of these observations will appear in *Nature* before Christmas. On the face of things, the circumstances are even odder than the *New York Times* suggests.

To anticipate what the authors of these discoveries have to say would be unfair to everybody concerned, readers not less than contributors, but just over a decade after the first discoveries of pulsating stars, it is clear that the study of these surprising objects is about to spring to life, and for several different reasons. Not least of all, the mere measurement of a pulsation rate of 642 Hz can be interpreted as the signal from a spinning neutron star only if it is supposed that the object is hovering on the edge of mechanical stability. If such a star has a mass roughly equal to that of the Sun, and a diameter of say 10 km as would be suggested by the equation of state for neutron matter, the velocity of the surface of the star would be a substantial fraction, say 10 per cent, of the speed of light, the rotational energy would be comparable with the energy supposed to be released in supernova explosions and the balance between gravitational attraction and rotational instability would be tenuous, to say the least. It is too soon to know whether some other model than that of the rotating solar-mass neutron star will be necessary to account for the observations now reported.

The other outstanding puzzle about the star itself is the

mechanism of its formation. The newly discovered pulsating star, thought to be at a distance of 2500 parsecs, has been radiating for at least twenty years and probably for much longer. It will be crucial to make accurate measurements of the rate of deceleration (reported by the *New York Times* to be negligible). Evidence for radiation elsewhere in the electromagnetic spectrum should also suggest what kinds of processes may have conspired towards the formation of this object (but it is reported that X-ray observations by the Einstein observatory have drawn a blank, which suggests the absence of a recently formed supernova shell). The possibility that the object may have been formed by some process not now considered feasible cannot be discounted, but as yet there is nothing to suggest what that might be.

Puzzles such as these apart, pulsating stars are plainly destined for closer attention because they have proved to be valuable pointers to the characteristics of the galactic environment. So much can be told from the careful measurement of the velocities of 26 pulsating stars measured by A.G. Lyne, B. Anderson and N.J. Salter with the radio-interferometer formed by the antennae at Jodrell Bank and at Defford, 127 kilometres to the south. The sample of radio objects includes those known before 1977, accessible from the Northern Hemisphere and such that a stable reference radio-source could be placed within the working aperture of the interferometer. (See *Mon. Not. Roy. Astron. Soc.*, November 1982.)

The results are important for being the most accurate measurement of the proper motions of pulsating stars, if otherwise unremarkable. So far as can be told from the two-dimensional projection of the velocity vectors on the sky, the motion of this sample of pulsating stars is isotropic. The distribution of speeds differs from a Maxwellian distribution in that there is an excess of high and low speeds — but then there is no reason why pulsating stars at their formation should acquire speeds characteristic of thermal equilibrium. A handful of the objects in the sample have speeds perpendicular to the plane of the galaxy that would be sufficient to carry them beyond the galaxy, while most of the others are destined to oscillate above and below the plane with periods measured in tens or even hundreds of millions of years.

The value of pulsating stars as guides to what the interstellar medium consists of seems also to be destined for fresh appraisal. From the earliest days the rotation of the plane of polarization from a distant radio source (the Faraday effect) has been recognised as a guide to average density of free electrons along the line of sight. Pulsating sources, invariably polarized, are especially convenient sources for this purpose. Now J.A. Roberts and J.G. Ables of the Radiophysics Division of the Australian Commonwealth Scientific and Industrial Research Organisation have formally published (*Mon. Not. Roy. Astron. Soc.*, **201**, 1119-38; 1982) their previously circulated account of how simultaneous records at several frequencies of the variation of the intensity of pulsating stars with time can throw light on the ways in which the radio-energy is scattered in the interplanetary medium. The technique is merely an elaboration of the scintillation measurement that are now customary, but the analysis suggests that it should be possible by a systematic effort to use pulsating stars as a means of keeping track of the gross features of the interstellar radio-scattering material. The only obvious snag is the patience that will be required — and the danger that observers will be dragged away from tasks like these by the discoveries that will succeed the 642 Hz pulsating star.

Self-splicing RNA

from John Abelson

THOMAS CECHE and his colleagues in Boulder, Colorado have been studying the transcription and processing of the ribosomal RNA in the protozoan *Tetrahymena thermophila* and have made a surprising and important discovery — the RNA appears to be able to catalyse the splicing out of its own intervening sequence.

In this species of *Tetrahymena* all of the macronuclear rRNA genes contain an 0.4 kilobase (kb) intervening sequence in the 26S rRNA coding region. Cech found that the excision of the intervening sequence from the precursor RNA, the circularization of the intervening sequence and very likely the covalent joining of the exon sequences are reactions which can take place *in vitro* and do not require either energy (in the form of ATP) or a protein¹.

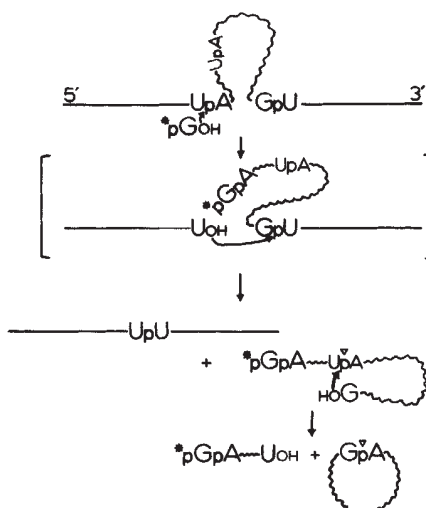
The conclusion that a protein is not required for the splicing reaction has come from studies on RNA obtained from *in vitro* transcription of a cloned fragment of the 26S rRNA gene from *Tetrahymena*. Earlier studies by Cech and colleagues² on the splicing of RNA transcribed in isolated *Tetrahymena* nuclei had suggested that the splicing reaction had unusual properties. The excision of the intervening sequence during *in vitro* transcription depends on monovalent cation concentration. In low salt, transcription of the rRNA genes is optimal and no intervening sequence is excised. At higher salt concentrations, the intervening sequence is released as a linear RNA molecule which is subsequently converted to a circle.

Using precursor RNA synthesized in low salt, the requirements for excision and circularization of the intervening sequence were found to be a monovalent cation (NH_4^+), a divalent cation (Mg^{2+}) and guanosine. The requirement for guanosine can be met by GTP, GDP, GMP or guanosine but not by 2'-deoxyguanosine, 2'-3'-dideoxyguanosine or 3'-GMP. Thus, there is a requirement for the 2'- and 3'-hydroxyls of guanosine but the hydrolysis of nucleotide is apparently not required for energy. When the precursor RNA was deproteinized by a variety of rigorous protocols including SDS-phenol extraction, boiling and protease treatment, it remained active in the excision reaction, requiring only salt, Mg^{2+} and guanosine.

If the G in the reaction is labelled, for example as ^{32}P -GTP, the label is found at the 5' end of the linear intervening sequence but not in the circular intervening sequence. During excision of the intervening sequence, the guanosine cofactor is linked to the intervening sequence in a 3'-5' phosphodiester bond³. The circularization of purified linear intervening sequence requires only added divalent cation and proceeds better at

elevated temperature. The products of the circularization reaction are a covalently closed circle that is 15 bases shorter than the linear intervening sequence and a linear fragment containing 15 bases of the 5' end.

These studies suggested that the ribosomal RNA was capable of self-splicing but the presence of a peptide with unusual properties was difficult to exclude. To deal with this, Cech and his colleagues¹ have now cloned a fragment of the *T. thermophila* 26S rRNA gene into a plasmid containing the *Escherichia coli* lac UV5 promoter. *In vitro* transcription of this DNA with purified *E. coli* RNA polymerase produced a transcript



The splicing reaction mechanism.

containing 35 bases of *lac* sequence, 261 bases of the 5' exon, the 413 base intervening sequence and 624 bases of the 3' exon. This transcript was deproteinized and was shown to undergo self-catalysed excision and circularization of the intervening sequence. In addition, the products of splicing were investigated by nuclease S_1 mapping using a DNA fragment lacking an intron. The results of the study were compatible with precise exon joining.

These results require a model that by necessity takes us into unfamiliar territory. We have an RNA that is apparently capable of a series of phosphotransfer reactions. The reactions are self-catalysed. The RNA is not an enzyme since there is no evidence that any RNA molecule can do more than catalyse self reaction. Cech and colleagues suggest the new term 'ribozyme'. Whether the reaction is catalysed by an enzyme *in vivo* or not remains an open question but there is no *a priori* requirement for an enzyme. The excision reaction is only 60 times faster *in vivo* than *in vitro* and this difference can easily be explained by factors present in the cell which might stabilize the RNA structure.

In the present working model of the splicing reaction, which will be described in detail in a paper soon to be published in *Nature*⁴, the splicing reaction is initiated by the binding of G to a site in the RNA in close proximity to the 5' splice junction (see the figure). The phosphodiester bond at the 5' splice junction is broken, perhaps due to strain provided by the tertiary structure of the RNA, and the 5' phosphate is transferred to G forming a 3'-5' phosphodiester bond. The 3' splice junction is similarly opened with transfer to the 3' hydroxyl of the 5' exon to form the spliced rRNA. The linear intervening sequence is released. In the subsequent circularization reaction a phosphodiester bond 15 nucleotides from the 5' end of the intervening sequence is opened. In this case, transfer is to the 3' end of the intervening sequence to form the circle with the release of the 5' oligonucleotide. In this mechanism, each of the three cleavage reactions is accompanied by a ligation reaction, that is, the process consists of a series of three phosphotransfer reactions. These reactions are expected to be reversible. Presumably they are driven in the direction of excision by the high concentration of guanosine and perhaps by changes in the structure of the RNA.

The unfamiliar aspect of this mechanism is, of course, that we must assume that the RNA is folded in such a way as to form a guanosine-binding site and to bring the phosphodiester bonds which are to be exchanged into very close proximity. Water must be excluded from the active site to prevent hydrolysis. There are other examples of phosphotransfer reactions which do not require energy, notably the DNA nicking-closing enzymes.

The mechanism of RNA splicing described here is, by its nature, likely to have been a primitive one. One might expect that all RNA splicing would proceed by a similar mechanism; but this is not the case. In the other system where detailed information is known of the mechanism — tRNA splicing in yeast — the reaction proceeds by a totally different mechanism. At least two enzymes catalysing as many as four reactions are required and the ligation step requires two molecules of ATP^{5,6}. Little is known about the mechanisms of mRNA splicing but it will not be surprising if that reaction proceeds by yet a third mechanism.

Since intervening sequences have only been found in the genes of eukaryotes, it has been assumed by some that splicing is a reaction which emerged rather late in evolution. The discovery of self-splicing RNA brings that assumption into question.

1. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. & Cech, T.R. *Cell* 31, 147 (1982).
2. Cech, T.R., Zaug, A.J. & Grabowski, P.J. *Cell* 27, 487 (1982).
3. Zaug, A.J. & Cech, T.R. *Nucleic Acids Res.* 10, 2823 (1982).
4. Zaug, A.J., Grabowski, P.J. & Cech, T.R. *Nature* (in the press).
5. Greer, et al. *Cell* (in the press).
6. Peebles, C.L. et al. *Cell* (in the press).

If an RNA molecule can have the intrinsic property of specifically breaking and rejoining its own bonds and of forming specific binding sites for nucleotides, certainly those properties would have been put to use in early evolution. As Cech and his colleagues point out, the reverse of the excision reaction is integration. This raises

the possibility of 'RNA transposons'. Shuffling of RNA sequences by these mechanisms could have been an important factor in the generation of sequence diversity in primordial organisms. □

John Abelson is in the Division of Biology, California Institute of Technology, Pasadena, California.

Inserted retroviruses cause embryonic lethal mutation

from Keith R. Willison

EXPERIMENTS in which genes and recombinant DNA molecules have been transferred heritably into the chromosomes of mice have resulted in the identification of a gene whose expression may prove to be essential for a specific stage of embryonic development. Although natural stage-specific embryonic lethal mutants are known, none of the mutant genes has yet been isolated and their mode of action has remained unclear. Rudolf Jaenisch has now used the alternative approach of interfering with the expression of a normal gene through the insertion of retroviral sequence into a mouse chromosome. This has enabled him to produce artificially an embryonic lethal mutation whose site of action he can locate with a retroviral probe.

This experimental approach and its possible applications have been discussed in earlier *News and Views* articles^{1,2}. Jaenisch decided, some ten years ago, to try to introduce novel genetic material into the mouse and study its effect on development by infecting early mouse embryos with tumour viruses. Both DNA and RNA tumour viruses were tested and it was quickly found that the Moloney murine leukaemia virus (M-MuLV) was the most promising in the sense that the viruses were able to become integrated in the DNA of the sperm and eggs⁴ and could be inherited by later generations of mice. This in itself was an important result at the time (before the advent of gene cloning) because it proved that such events could be made to occur in the laboratory as well in the wild. That this does indeed happen in natural environments has been amply confirmed in cats⁵, mice^{3,6} and chickens⁷. Some of the mice which carry M-MuLV sequences eventually express them in lymphatic tissues, become full of virus and develop leukaemia.

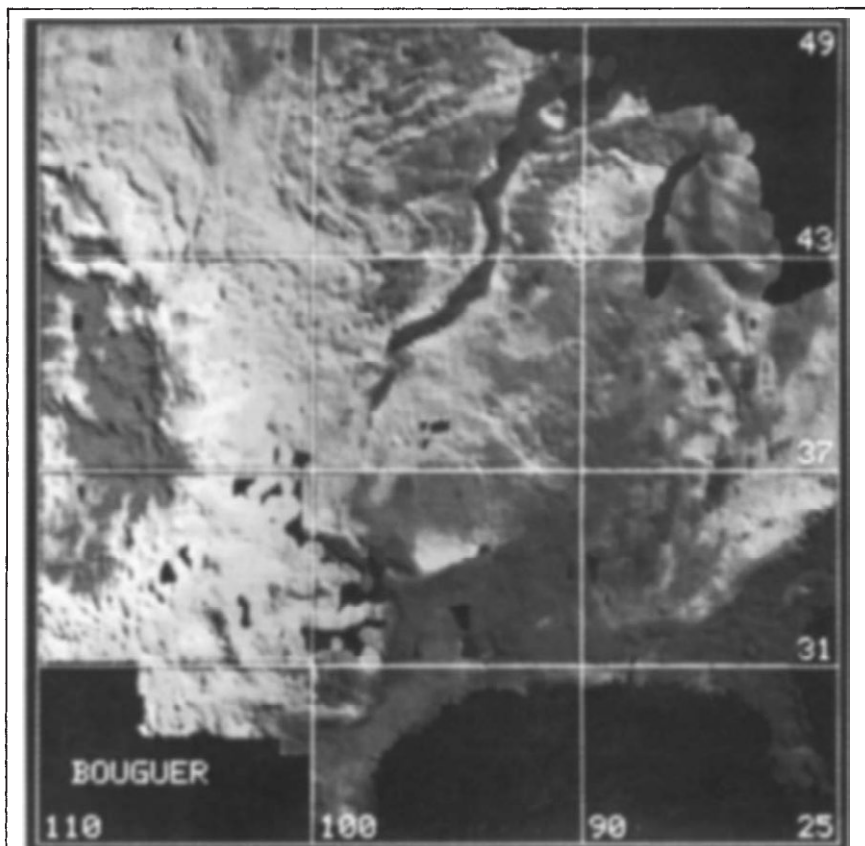
Jaenisch has constructed 13 mouse strains containing integrated M-MuLV sequences called *Mov 1-13*. They were produced by infecting embryos at different stages with virus or purified DNA using microinjection techniques^{4,8,9,10} (see also

ref. 1 for more detail). The strains are maintained as heterozygous stocks in order to make it easier to breed healthier animals, and the M-MuLV-containing chromosome is followed either by measuring viraemia in the offspring or by Southern blotting with purified M-MuLV probes. The latter technique allows one to distinguish between animals homozygous or heterozygous for the *Mov* allele in cases where the genomic flanking sequences have been cloned, and therefore aids breeding for *Mov*-homozygous mice.

The construction of homozygous strains was described by Jaenisch at a recent meeting* of the Ciba foundation in London and has given an important result. The attempt to produce *Mov-13* homozygotes resulted in consistently smaller than expected litter sizes for the particular inbred genetic background on which it was carried (C57BL/6). This suggested that some of the embryos might be dying, that is, that the *Mov* allele was behaving as an embryonic lethal gene. In fact it turned out that the homozygous embryos (*Mov-13/Mov-13*) were dying at around day 12 of gestation. There appeared to be no specific malformations but rather a total degeneration of the embryo. As the DNA does not quickly break down in dying embryos, it could be isolated and analysed to show that the degenerating embryos are indeed homozygous for *Mov-13*. The lethality of *Mov-13* homozygotes was stage-specific and absolute; 55 embryos surviving beyond day 14 were analysed and none was homozygous for *Mov-13*.

Are there precedents for the production of mutations by retroviruses? It is worth noting that a natural integration of an

*The Ciba Foundation meeting on the 'Molecular Biology of Oogenesis' was held 9-12 November. The proceedings will be published as Ciba Symposium 98 (Pitman Books).



Digitally processed gravity measurements have revealed the presence of an ancient rift structure extending across Missouri. In this image the rift extends from the top (centre right) to the middle (centre left). The structure, 120 to 160 km wide and 700 km long, is not discernible from surface topography. E.A. Guinness *et al.* (*J. geophys. Res.* **87**, 8529; 1982) identify the feature as the signature of a Precambrian rifting event that failed to develop into a spreading zone.

Keith R. Willison is in the Chester Beatty Research Institute, Fulham Road, London SW3 6JB.

MuLV provirus has also given rise to the dilute (*d*) coat colour mutation of DBA/2J mice¹¹, and in tissue culture Varmus¹³ has used M-MuLV as an insertion mutagen to isolate revertants of a rat cell line transformed by Rous sarcoma virus. In this latter case it was directly shown that the M-MuLV had inserted into the resident RSV provirus sequence in the rat genome and affected the expression of RSV RNA species.

How does the *Mov-13* allele cause embryonic death? Is it just toxicity due to viraemia? No virus can be detected in heterozygotes until day 16 of development and viraemia would certainly be expected to be dominant. Has the M-MuLV DNA integrated into a gene or gene complex essential for embryonic development? It has certainly inserted into actively transcribed DNA, as Jaenisch also reported at the meeting. Furthermore, transcription of DNA at this site normally begins at day 12 of gestation when in normal embryos two abundant transcripts of 4kb and 5kb can be detected in the total RNA. The level of expression of these two transcripts changes during development, maximum expression occurring between days 14 and 18 when at

least 10-fold more 4kb and 5kb RNAs can be detected than on either day 12 or day 19 of development. There appears to be no expression of these transcripts in adult tissues, although very small quantities of a 7–8kb transcript are detectable. The relationship to the embryonic 4–5kb RNAs is not yet known. A reasonable interpretation of these data is that an embryonic lethal mutation has been produced by insertion mutagenesis. The affected gene appears to be expressed primarily during embryogenesis and its identification and cell distribution will be awaited with great interest. □

1. Hogan, B. & Williams, H. *Nature* **294**, 9 (1981).
2. Petri, W. *Nature* **299**, 399 (1982).
3. Jaenisch, R. & Mintz, B. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1250 (1974).
4. Jaenisch, R. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1260 (1976).
5. Benevise, R.E. & Todaro, G.J. *Nature* **252**, 456 (1974).
6. Cohen, J.C. & Varmus, H.E. *Nature* **278**, 418 (1979).
7. Frisby, D., Weiss, R.A., Roussel, M. & Stehelin, D. *Cell* **17**, 623 (1979).
8. Harbers, K., Jahner, D. & Jaenisch, R. *Nature* **293**, 540 (1981).
9. Jahner, D. & Jaenisch, R. *Nature* **287**, 456 (1980).
10. Jaenisch, R., Jahner, D., Nosib, P., Simon, I., Lohler, J., Harbers, K. & Grotkopp, D. *Cell* **24**, 519 (1981).
11. Jenkins, N., Copeland, N.G., Taylor, B.A. & Lee, B.K. *Nature* **293**, 370 (1981).
12. Varmus, H., Quintrell, N. & Ortiz, S. *Cell* **25**, 23 (1981).

that produce significant amounts (~40–100 per cent) of the mRNA. Sequencing of the cloned genes from Hikone-R and Seto flies confirms a deletion (marked ORL on the figure) from –305 to –356 (replaced by the nucleotides CGC) relative to Oregon-R. In the Berkeley-1 (BER-1) strain no *Sgs 4* mRNA has been detected at any larval, pupal or adult stage tested. Here, sequencing of the cloned gene confirms a 95 bp deletion from –392 to –486 (replaced by GAT). Both producer and nonproducer strains have the same TATA box at –22 to –28 and essentially the same loose match with the consensus sequence 5'-GG⁵CCAATCT-3' at –66 to –74 (the only change being the substitution of a C for the 3'-terminal T in BER-1).

In nuclei from embryos or tissue culture cells, three DH sites can be observed 3' to the *Sgs4* gene but none near the 5' end. The 3' sites persist in the chromatin of the third-instar larval salivary gland, but in addition major sites can be mapped to +30, –70, –330, –405 and –480 (with an estimated accuracy of ±30–50 bp) relative to the start of transcription. The most prominent site is that centred at –405 (mapped within ±5 bp). The site mapped to –330, which is lost in the Oriental strains, might play a part in determining the efficiency of transcription. The other DH sites are still apparent in the chromatin of these larvae at the same DNA sequences⁸. In contrast, all the tissue-specific 5' DH sites are lost in the nonproducer BER-1 strain⁸, even though there are no major differences observed in the DNA sequence other than the deletion¹⁰. In particular, the sequences from –390 to –250 are identical for Oregon-R and BER-1 flies. This indicates that the DNA sequences at or near the major DH site have a role in determining the chromatin structure of the surrounding region.

That this function may reflect a switch in the underlying DNA structure is suggested by the work on chick globin genes described above. It must be noted, however, that not all sites which are *S₁* sensitive in supercoiled plasmids correspond to known DH sites in chromatin¹¹. Conversely, it seems likely that not all chromatin DH sites will be *S₁* sensitive. Nonetheless, it is apparent that in some instances, specific DNA sequences near the 5' ends of globin genes can adopt different conformations dependent on supercoiling of the DNA. That such a switch occurs *in vivo* is indicated by the observation that *S₁* nuclease also detects sensitive sites in chromatin, in the region of the previously characterized tissue-specific and developmentally specific DH sites¹¹. The same sites are recognized by reagents such as bromoacetaldehyde, which react with unpaired regions of DNA (T. Kohwi and H. Weintraub, personal communication).

What might be the nature of this change? What sequences might be involved? Certainly the most radical suggestion would be

Chromatin structure, DNA structure

from Sarah C.R. Elgin

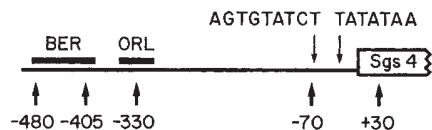
DURING the past few years, specific features of eukaryotic chromatin structure, in particular sites hypersensitive to digestion by DNase I, have been detected and mapped for a variety of genes^{1–4}. Frequently (but not always) such sites have been observed at or near the 5' ends of genes transcribed by polymerases I and II (see ref. 5 for a brief review). Several studies have noted that 5' DNase I-hypersensitive (DH) sites are present before transcription begins, suggesting they may have some functional role^{2,4,6,7}. Now, a number of recent studies is providing evidence that DH sites can play a critical part in the control of gene expression.

A cluster of five DH sites, found only in nuclei where the gene is expressed, has been mapped upstream from the *Drosophila melanogaster* glue protein gene *Sgs4* (ref. 8). In a strain that does not express *Sgs4* (ref. 9), the DH sites are not observed⁸ and sequence data reveal that a small upstream deletion has removed the major DH site of the complex¹⁰ — implying that the loss of this segment both affects the chromatin structure of the surrounding region⁸ and eliminates gene expression. A clue to a possible underlying mechanism comes from work on the chick globin locus¹¹. Here DNA sequences at or near the previously mapped chromatin DH sites can

adopt (at least) two different conformations. When relevant DNA segments are subcloned into pBR322 and the recombinant plasmids are digested with *S₁* nuclease, these sites are cleaved in the negatively supercoiled DNA molecule, but not in the linear DNA molecule¹¹. We may find that there are switches relevant to gene expression constructed from DNA sequences that can possess alternative structural configurations, the different forms being stabilized and amplified in the chromatin structure. The major DH site of the *Sgs4* locus is a good candidate for such a switch.

The *Sgs4* glue gene of *D. melanogaster* is normally expressed in the salivary gland of the third-instar larva. Differences in mRNA production, observed for different strains, correlate with the presence of small upstream deletions^{9,10}. Three Oriental strains (Seto, Hikone-R and Kochi-R) that produce only trace amounts (~2 per cent of levels observed in Oregon-R) of the *Sgs4* mRNA have an upstream restriction fragment that is shorter by about 50 base pairs (bp) than that of five strains (Oregon-R, Chiete V, Urbana-S, Canton-S and D323)

Sarah C.R. Elgin is in the Department of Biology, Washington University, St Louis, Mo 63130.



Map of the 5' end of *Sgs4* from *D. melanogaster* Oregon-R. DH sites are indicated by broad arrows. The positions of the control sequences and deletions discussed are indicated.

a switch in the form of the DNA double helix, for example from B to Z. This transition can occur in physiological conditions in a segment of d(pCpG)_n, the Z form being stabilized by the introduction of negative supercoiling^{12,13}. It seems likely that Z-DNA could also be stabilized by protein binding¹⁴. The B-Z boundary is sensitive to S₁ nuclease¹³.

However, a less radical shift in DNA structure seems more likely. No common features of sequence organization have so far been identified for DH sites. It has also been reported that enhancer elements of similar function, such as those of SV40 and of polyoma virus, do not show any obvious sequence homology¹⁵. It is possible that recognition of such sites will come only with a greater appreciation of the three-dimensional structure of the DNA. That the conformation of the DNA double helix, including the structure of the sugar-phosphate backbone, is variable and sequence dependent, even within the overall B form, is clear both from recent studies of the helical periodicity^{16,17} of B-DNA and from the X-ray crystallographic analysis of the B-DNA dodecamer CGCGAATTCGCG¹⁸. The sequence-dependent conformation is apparently reflected in the sensitivity of the DNA to endonucleases, there being a good correlation between the sites of rapid DNase I cutting in this dodecamer and

positions of high local helical twist¹⁹. The rate of cleavage at a particular site can reflect not only the immediate site but also the 'DNA environment', as illustrated by the fact that the four different *Pst*I sites in pSM1 are cut at very different rates²⁰. This effect is not dependent on the degree of supercoiling. It seems, then, that it may be possible to obtain empirical maps of DNA structure by examining site-specificity in the interaction of DNA with enzymes (such as micrococcal nuclease)²¹ or with various

small DNA-binding compounds (for example intercalators)^{22,23}. Structural patterns in the DNA that are invariant to supercoiling might be related to relatively stable features of chromatin structure, while those that are altered might be candidates for potential switch regions as discussed above. Clearly a more detailed appreciation of DNA structure is required to understand the code that directs protein-DNA interactions in eukaryotic chromatin. □

Chromosomes and cancer

from David Forman and Janet Rowley

THE analysis of chromosomes in relation to cancer is a field of research which has gone through successive periods of excitement and frustration. The enthusiasm in the early 1960s arising from the discovery that chronic myeloid leukaemia is strongly associated with a specific marker (the Philadelphia chromosome Ph¹; see also Logan, J. & Cairns, J. *Nature* **300**, 104; 1982) diminished when it was realized that few other cancers exhibit such clear-cut abnormalities. Since then there have been critical developments in technique which have brought the chromosome back to the centre of cancer science.

These technical developments have been of two types. First came consistent improvements in the banding of human chromosomes, which revealed several translocations and deletions, beginning with the discovery that the Ph¹ chromosome results from a translocation usually involving chromosomes 9 and 22. Second, and more recently, has come the application of recombinant DNA technology to the human genome. As was emphasized at a recent meeting on chromosomes and cancer*, the result of these changes has been a narrowing of the gap between molecular biology, dealing with units of around 10³ nucleotides upwards, and cytogenetics, dealing with around 10⁶ nucleotides downwards. Although the gap is not yet closed, communication between the two fields is becoming intense and vigorous. Not only are more cancers being related to consistent non-random chromosome changes but important clues are also being found as to how these alterations affect function.

The most exciting advances reported at the meeting came from Croce (Wistar Institute) and Taub and Leder (Harvard Medical School) concerning Burkitt's lymphoma — already known to be closely associated with an (8;14) translocation.

Chromosome 14 contains the immunoglobulin heavy-chain gene cluster which is active in Burkitt's lymphoma cells. Both groups showed that the break in chromosome 14 occurs within the cluster and that the break in chromosome 8 is at the site of the *c-myc* oncogene. Croce and co-workers at the Wistar Institute found the heavy-chain gene regions from several independently derived Burkitt's lymphoma cell lines hybridized with *c-myc* probes. The Harvard group, using a probe derived from a mouse plasmacytoma, isolated a homologous segment of DNA from 12 Burkitt's lymphoma cell lines. In two cases a single restriction fragment contained both a portion of the heavy-chain gene and the cellular *c-myc* gene. Both groups found that about a third of all lines tested contained rearrangements of the immunoglobulin heavy-chain region caused by the integration of the *c-myc* region. Croce suggested that in these cases the genes are aligned in a 5' to 5' head-to-head fashion. The suggestion is therefore that the level of expression of the transforming *c-myc* gene after translocation is affected by the neighbouring active immunoglobulin genes and normal regulatory control is lost. It is not without significance that the other chromosomes involved with chromosome 8 in translocation events in Burkitt's lymphoma are numbers 2 and 22, containing the sites for the immunoglobulin κ and λ light-chain genes respectively, and that the lymphomas express κ chains in the case of (8;2) translocations and λ in the case of (8;22) translocations.

An animal B-cell tumour, the mouse plasmacytoma, also involves an homologous translocation — a (12;15) or occasionally a (6;15) translocation (Spira, Karolinska Institute). Again chromosome 12 contains the mouse immunoglobulin heavy-chain gene and chromosome 6 the κ light-chain gene. The *c-myc* gene is known to be located on chromosome 15 in mouse

David Forman is in the ICRF Cancer Epidemiology Unit, The Radcliffe Infirmary, Oxford OX2 6HF and Janet Rowley in the Section of Haematology/Oncology, University of Chicago, Illinois 60628.

*The fifth Bristol-Myers symposium on cancer research — 'Chromosomes and Cancer: From Molecular to Man' — was held at the University of Chicago on 18–19 October 1982.

1. Wu, et al. *Cell* **16**, 797 (1979).
2. Wu, C. *Nature* **286**, 854 (1980).
3. Stalder, J. et al. *Cell* **20**, 451 (1980).
4. Keene, M.A. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 143 (1981).
5. Elgin, S.C.R. *Cell* **27**, 413 (1981).
6. Weintraub, H., Beug, H., Groudine, M. & Graf, T. *Cell* **28**, 931 (1982).
7. Lowenhaupt, K., Cartwright, I.L., Zimmerman, J.L., Keene, M.A. & Elgin, S.C.R. (submitted).
8. Shermoen, A.W. & Beckendorf, S.K. *Cell* **29**, 601 (1982).
9. Muskavitch, M.A.T. & Hogness, D.S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7362 (1980).
10. Muskavitch, M.A.T. & Hogness, D.S. *Cell* **29**, 1041 (1982).
11. Larsen, A. & Weintraub, H. *Cell* **29**, 609 (1982).
12. Peck, L.J., Nordheim, A., Rich, A. & Wang, J.C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4560 (1982).
13. Singleton, C.K., Klysik, J., Stirdivant, S.M. & Wells, R.D. *Nature* **299**, 312 (1982).
14. Lafer, E.M. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3546 (1981).
15. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299 (1981).
16. Peck, L.J. & Wang, J.C. *Nature* **292**, 375 (1981).
17. Rhodes, D. & Klug, A. *Nature* **292**, 378 (1981).
18. Dickerson, R.E. & Drew, H.R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7318 (1981).
19. Lomonosoff, G.P., Butler, P.J.G. & Klug, A. *J. molec. Biol.* **149** 745 (1981).
20. Armstrong, K. & Bauer, W.R. *Nucleic Acids Res.* **10**, 993 (1982).
21. Keene, M.A. & Elgin, S.C.R. *Cell*, **27**, 57 (1981).
22. Cartwright, I.L. & Elgin, S.C.R. *Nucleic Acids Res.* (in the press).
23. Jessee, B., Gargiullo, G. & Worcel, A. *Nucleic Acids Res.* (in the press).

and several investigators, as well as the group at Harvard, have shown that it is translocated adjacent to the immunoglobulin locus.

If *c-onc* genes in aberrant positions are critical to tumour development, what is their normal function? Lane and Cooper (Sidney Farber Cancer Institute) have identified five different stage-specific transforming genes active in lymphocytic leukaemias and lymphomas. Tumours which share a common phenotype, representative of a discrete stage in normal lymphoid differentiation, appear also to share a common transforming gene, suggesting that the altered genes may normally be involved in growth regulation during cellular differentiation. The investigators have gone on to examine a series of African and American Burkitt's lymphomas by transfection of NIH 3T3 cells and have identified a common transforming gene activated in these tumours which is not analogous to genes from the Epstein-Barr virus genome or indeed to *c-myc*. Thus different genes may be involved at different stages of the neoplastic process and in the world outside NIH 3T3 cells, tumour development may depend on the activity of more than a single oncogene.

In this respect it is interesting that the oncogene isolated from a bladder carcinoma cell line and which is homologous to the *ras*^H gene maps to the exact region of chromosome 11 that is deleted in the Wilms tumour-aniridia syndrome (11 p13; Francke, Yale University). Thus here there is an association between one particular oncogene and two quite different tumours. It is also apparent that particular translocations may be associated with variable but related tumour phenotypes. The (9;22) translocation is involved in certain cases of acute lymphocytic and non-lymphocytic leukaemias as well as in established chronic myeloid leukaemia. High-resolution banding techniques have shown that the break point on chromosome 22 is very similar in all three diseases (Yunis, University of Minnesota). If there were an oncogene associated with either chromosome 9 or chromosome 22, additional factors must determine which type of phenotype is expressed.

In general, these new advances emphasize the contribution made by cytogeneticists in providing an expanding data base which allows molecular biologists to decide which areas of the genome to concentrate on. Yunis and co-workers can now distinguish several thousand chromosome bands in the human genome so that previously masked small chromosome defects can now be visualized. Certainly it is clear that chromosomal abnormalities are the rule rather than the exception in haematological malignant diseases, with over 90 per cent of all types of acute leukaemia exhibiting some form of

defect.

A computerized cancer chromosome registry has now been established (Metelman, University of Lund) and is revealing that although most leukaemias are associated with several types of defect, there are distinct patterns in these associations. First, the abnormalities usually only involve a limited number of chromosomes. In the case of acute myeloid leukaemia, although there is not an association with one specific change, there is a strongly preferential involvement of chromosomes 5, 7, 8, 15, 17 and 21. Second, it is usually the case that specific segments of the chromosomes are involved in the aberrations (for example, bands q33-36 on chromosome 7). Third, detailed analyses are making possible the beginnings of a cancer chromosome epidemiology. For example, the incidence or frequency of chromosome changes associated with acute non-lymphocytic leukaemia (ANLL) that arise from chemotherapy or irradiation following a previous malignancy are different from the changes that take place when the disease occurs *de novo* or in childhood (Rowley, University of Chicago). In childhood an (8;21) translocation or an 11q deletion is often seen in ANLL. With increasing age and with exposure to chemicals, either occupationally or as a result of cytotoxic therapy, deletions of part or all of

chromosomes 5 or 7 become much more prevalent and the (8;21) or 11q effects very rare. The specific type and degree of abnormality can often have clinical implications in terms of both diagnosis and prognosis (Golumb, University of Chicago).

Further technical developments will soon be ready to be profitably employed in research into chromosome neoplasia. Flow cytometry can now individually separate most human chromosomes and has already been used for mapping studies (Carrano, Lawrence Livermore, National Labs) and the preparations of single chromosome DNA libraries (Young, Beatson Research Institute). Ward (Yale University) described the *in situ* hybridization of as little as 20-30 pg of biotin-labelled DNA to the chromosome with subsequent detection and visualization using the protein avidin and alkaline phosphatase. It should soon be possible to rapidly screen individual chromosomes with probes for specific genes and build up a much more detailed account of the structure of the chromosome in neoplastic cells.

The increasing fund of knowledge at both the cytological and molecular levels, together with the arrival of new techniques, should lead to rapid advances in our understanding of the role of chromosomes in cancer and, hopefully, to much improved therapy. □

Sex reversal and sex determination

from Elizabeth Simpson

THAT mammals having the XX sex chromosome constitution are generally female, while those with XY constitution are generally male has often been interpreted as showing that a gene (or genes) on the Y chromosome controls primary sex determination. When a Y chromosome is present the foetal gonad, which has the potential to become either ovary or testis, differentiates into a testes and, when absent, into an ovary — the first and crucial event in a series of changes that determines whether the organism has a male or female phenotype. Recent genetic analysis shows that this first step is not as simple as first thought and appears to be controlled by an interaction between autosomal genes and a gene(s) on the Y chromosome, for normal male differentiation.

Insight into sex determination has been provided by exceptions to the rule that the presence of the Y chromosome leads to a male phenotype. In mice, which provide the most readily examined examples, sex

reversal can be partial or complete, and is characterised by XX individuals that have testes (rarely ovotestes) and are apparently normal phenotypic males, and XY individuals that have ovotestes or ovaries, and are apparently normal or near normal phenotypic females.

The *Sxr* (sex reversed) gene, originally described by Cattanaach, causes XX mice inheriting it to develop as phenotypic males. It was at first thought to be a dominant, autosomal gene, probably derived by translocation from the Y-chromosome. However, it proved impossible to map the *Sxr* gene to any of the autosomes — a mystery that has only recently been solved by the finding that *Sxr* is located on the distal arm of the X chromosome in XX sex-reversed mice.

The mapping of the *Sxr* gene on the X chromosome comes from two independent lines of evidence. First, experiments described in this issue of *Nature* by McLaren and Monk (p.446) and Cattanaach and colleagues (p.443 & p.445) show that some *Sxr* XX mice develop as females or hermaphrodites rather than males if the

Elizabeth Simpson is in the Clinical Research Centre, Transplantation Biology Section, Harrow HA1 3UJ.

paternally inherited X chromosome is preferentially inactivated (paternal X inactivation was ensured by introducing Searle's translocation (T[X;16]16H³) through the mother). Second, *Sxr* has been mapped to the X chromosome by Singh and Jones⁴ using hybridization *in situ* of Y chromosome specific DNA sequences (Bkm) to the distal end of the X chromosome in XX *Sxr* mice. From these findings Eicher⁵ and Burgoyne⁶ have gone on to propose that the Y chromosome contains two regions of Bkm positive sequences, one in the normal centromeric position and the other at the distal end of the chromosome. This makes it possible to explain why sex reversal is inherited in a pattern characteristic of an autosomal dominant gene, with half the XX progeny being sex reversed and half the XY progeny carrying the *Sxr* gene. During meiosis, when haploid sperm are generated by a carrier male, crossing over between X and Y chromosomes creates four types of sperm (i) an X carrying the translocated Bkm sequences from Y, (ii) a normal X, (iii) Y carrying both Bkm sequences, and (iv) a normal Y; created by the transfer of the distal Bkm sequence bearing end of the carrier Y to the X. Assuming that each sperm type has an equal chance to fertilize ova, this would produce the observed 1:1:1:1 ratio of *Sxr* XX males, XX females, XY *Sxr* carrier males, and XY noncarrier males. This interpretation of the transmission of *Sxr* is supported by recent cytogenetic data (Evans, Burtenshaw & Cattanaach, this issue of *Nature*, p.443).

Burgoyne⁶ has also argued that there may be a whole class of 'pseudo-autosomally' inherited genes located distal to the crossover point on the pairing segment of the X and Y chromosomes. These genes normally escape X inactivation and like *Sxr* would show the same pattern of apparently autosomal inheritance. Some of these genes, in addition to the normal Y specific Bkm related sequences present in the non-pairing part of the Y chromosome, may be involved in sex determination.

Recent work of Eicher and her colleagues⁵ with a series of XY sex reversed mice have indicated that the testis determining gene *Tdy* which is probably closely linked to the Bkm sequences, and maps near the centromere on the Y chromosome, cannot be the only primary sex determining gene: it needs to interact with other genes, perhaps in appropriate sequence for normal sex determination. One set of Eicher's sex reversed mice show inheritance of abnormal sex determination through a Y chromosome derived from *Mus poschiavinus*, a subspecies of *Mus domesticus*. An inbred strain of mice with this Y chromosome, designated Y^{POS}, breeds normally, and produces normal F₁ hybrids when mated with the C57BL/6 laboratory strain. However, when these hybrid males are backcrossed with C57BL/6 females half of the XY progeny are hermaphrodite (one gonad being an ovotestis). In subsequent backcross

generations the proportion of abnormal XY individuals increases until a high proportion of them are completely sex reversed, phenotypic females with two ovaries, and the rest were males with ovotestes, most of which are fertile⁷.

That there is a 1:1 ratio between males and hermaphrodites among the XY progeny of the first backcross generation argues for an effect of a dominant autosomal or pseudoautosomal gene (named *Tda-1*⁵) in addition to the abnormal testis determining gene *Tdy* transmitted by Y^{POS}. The increasing incidence of complete sex reversal in XY individuals in sequential backcross generations argues for the involvement of additional autosomal gene(s). Thus it would appear that in the original Y^{POS} carrying strain, the testis determining gene, *Tdy*, on Y^{POS} could interact normally with a dominant autosomal gene, *Tda-1*, and with additional autosomal gene(s) perhaps in carefully timed sequence, to determine primary sex (testis or ovary). However, when *Tdy* POS was isolated by backcrossing onto the C57BL/6 inbred genome, it could not interact normally with the corresponding genes of C57BL/6.

Three other examples of the type of sex reversal seen with Y^{POS}, have now been found in wild mice captured in Northern, Central and Southern Italy respectively. Each is inherited via the Y chromosome of a male *Mus domesticus*, and produces XY hermaphrodites on back-crossing to C57BL/6⁷.

That autosomal genes are involved in primary sex determination is also shown by the effects of an inherited dominant autosomal gene on chromosome 17, causing partial or complete sex reversal in XY individuals (Washburn and Eicher, in preparation; quoted in Eicher⁵).

Sex reversal may also be brought about by another Y-chromosome abnormality; an apparent rearrangement of the Y chromosome that leads in each generation to translocation of *Tdy* from the Y to the X chromosome during meiosis, in a manner

probably analogous to that of *Sxr* transmission. However, in this case the translocation involves a very large piece of the Y, readily visible on karyotyping. XX individuals carrying one such translocated X chromosome are phenotypic males, whilst the reciprocal translocation gives rise to females with one normal X and a tiny translocation product lacking the *Tdy* gene⁵.

It would appear that normal primary sex determination involves the *Tdy* gene located on the Y chromosome near the centromere, interacting with autosomal and perhaps pseudoautosomal (that is, X/Y linked) genes. Problems can arise through translocations of *Tdy* to other sites on the Y chromosome, especially to the distal sites which pair with the X chromosome during meiosis, since then transfer to the X chromosome is possible. *Sxr* and the translocation described by Eicher (see previous paragraph) probably arose in this way, and both result in the production of a normal X and an X that bears part of the Y. The autosomal genes that necessarily act in concert with *Tdy* also make possible abnormal sex determination due to improper timing of events which take place in the bipotential gonad (see Eicher⁵). Mutations, deletions and inappropriate alleles of such autosomal (and/or pseudoautosomal) genes might be involved.

It is still an open question as to how *Tdy* operates in this sequence of events, and whether its gene product is the male specific antigen which has been proposed as the differentiation antigen that causes testis differentiation. To complicate the matter further, it appears that there are two male specific antigens, one identified by T cells (the transplantation antigen, H-Y) and the serologically detected male (SDM) antigen. These two have been separated by testing mice with the sex chromosome constitution XO. These females, although H-Y negative⁸ are SDM positive (ref 9, Koo, G. and Müller, U., personal communications). It has been argued¹⁰ that neither of the two male specific antigens are candidates for the role of sole determinant of sex differentiation, since XO female mice are SDM positive but have ovaries, whilst XY^{POS} females are H-Y positive but have ovaries. The H-Y typing, using skin grafting and/or cytotoxic T cells, of a variety of mice whose karyotype is abnormal or unexpected, for example, XO (H-Y negative, female⁸), XX *Sxr* (H-Y positive, male^{11,12}), XY^{POS} (H-Y positive, female¹³) is consistent with H-Y being tightly linked to *Tdy*, which is clearly not the sole testis determining gene. SDM typing is more problematical: from the XO data⁹ the structural gene for SDM cannot be on the Y chromosome and is not *Tdy*. Moreover, SDM typing is difficult to reproduce and none of the SDM monoclonals tested so far have shown male specific binding (see ref. 14 and Simpson & Goodfellow, unpublished).

By other less direct tests they are reported to have male specific activity¹⁵,

1. Ohno, S. *et al.* Gross, F. (eds) in *Monogr. Endocr.* Vol.2 Springer, New York, 1979).
2. Cattanaach, B.M., Pollard, C.E. & Hawkes, S.G. *Cytogenetics* 10, 318-337 (1971).
3. Lyon, M.J. *et al.* *Cytogenetics* 3, 306 (1964).
4. Singh, L. & Jones, K.W. *Cell* 28, 205-216 (1982).
5. Eicher, E.M. in *Prospect For Sexing Mammalian Sperm* (Colorado Associated University Press, Boulder, Colorado, 1982).
6. Burgoyne, P. *Hum. Genet.* (in the press).
7. Eicher, E.M., Washburn, L.L., Whitney, J.B. & Morrow, K.E. *Science* (in the press).
8. Simpson, E., McLaren, A. & Chandler, P.R. *Immunogenetics* (in the press).
9. Engel, W., Klemme, B. & Ebrecht, A. *Hum. Genet.* 57, 68 (1981).
10. Silvers, W.K., Gasser, D.L. & Eicher, E.M. *Cell*, 28, 439 (1982).
11. Bennett, D. *et al.* *Nature* 265, 255 (1977).
12. Simpson, E. *et al.* *Immunogenetics* 13, 355 (1981).
13. Johnson, L.L. *et al.* *Devel Genet.* (in the press.)
14. Gore-Langton, R.E., Tung, P.S. & Fritz, I.G. *Cell* (in the press).
15. Koo, G. *et al.* *Hum. Genet.* 57, 64 (1981).
16. Goodfellow, P.N. & Andrews, P.W. *Nature* 295, 11 (1982).

although at titers many orders of magnitude lower than that shown by monoclonals to other cell surface determinants that readily stimulate antibody production (for example, H-2, Thy-1, Ly antigens)¹⁶.

In conclusion then, it appears that the differentiation of testis from the bipoten-

tial gonad, the key step that leads to the development of the male phenotype, is controlled by an interaction between autosomal genes and genes on the sex chromosome. Whether sex specific antigens are involved, and if so, how, is still a matter of speculation. □

mixed with a small amount of water frost.

Stretching the capabilities of the size determination measurement to the outer Solar System, Morrison, Cruikshank and Brown have placed significant upper limits on the sizes of Triton, the major satellite of Neptune, and of Pluto. The results are consistent with icy objects of sub-lunar size, and with the hypothesis that the two objects are physically similar. The possibility that Triton is the largest planetary satellite can now be excluded as it is certainly smaller than the giant satellites Ganymede, Callisto and Titan. An improvement of only a factor of two in the quality of the IR measurements is likely to lead to positive detections, and hence definitive size measurements, for both Triton and Pluto.

The scientific value and cost-effectiveness of these Solar System studies are self-evident, and it is gratifying that NASA has reversed its earlier decision to stop funding for the Infrared Telescope Facility. The importance of IR astronomy for targets outside the Solar System is also gradually being realized in the study of star formation, galactic nuclei and many other dusty interstellar environments. It is to be hoped that the facility will long continue to exploit the special advantages of Mauna Kea, where, as well as University of Hawaii telescopes, there are now two international observatories in operation; the Canada-France-Hawaii Telescope (CFHT) and the United Kingdom Infrared Telescope (UKIRT). More telescopes are planned for the future, including the SERC Millimetre Wave Telescope. □

Distant planetary satellites more accurately measured

from Ian Gatley

A BASIC step in the exploration of the Solar System is the determination of the sizes of planets, satellites and asteroids. Most of these objects are far too small for direct-imaging techniques from ground-based telescopes, but the cost of spacecraft makes size determinations from the ground extremely desirable. Fortunately, a simple and elegant method which combines the results of optical and IR observations is available, and two papers in this issue of *Nature* give new results for four of the uranian satellites (Uranus is to be visited by Voyager 2 in 1986), Neptune's largest satellite Triton and the most distant planet, Pluto.

The principle is this. Sunlight falling on the target is either absorbed or reflected; the energy absorbed is subsequently re-radiated in the thermal IR. Visual photometry measures the reflected component of the solar radiation and IR radiometry measures the absorbed component and the sum of these is equal to the total incident energy. One can therefore solve simultaneously for the albedo (reflectivity) and the size of the object under study (Morrison *Icarus* 19, 1; 1973).

The principle is simple but its application is not, for two reasons: the IR signals are weak, and the atmosphere transmits poorly at the relevant IR wavelengths (around 20 μm) because of the presence of water vapour. These problems can be overcome by the use of a large IR telescope on a high dry site, such as the newly commissioned 3 metre NASA Infrared Telescope Facility on the 14,000 foot summit of Mauna Kea in Hawaii. This telescope was constructed primarily for the study of the Solar System, and its early results in this area are fascinating.

R. Hamilton Brown, Dale Cruikshank and David Morrison have measured the diameters and albedos of four satellites of Uranus — Ariel, Umbriel, Titania and Oberon, and find lower albedos and larger diameters than previous, less reliable estimates. Having made accurate size measurements, these workers now stress the importance of experiments to deter-

mine masses, for knowledge of the densities of the satellites will provide important insights into the origin and subsequent evolution of the uranian system. The albedo measurements give information on the surface composition of the satellites. Ariel, Umbriel, Titania and Oberon all have sizes similar to the large icy saturnian satellites, particularly Rhea, but they have significantly lower albedos, more like those found for Hyperion. It is noteworthy that Voyager images of Hyperion show it to be a reddish colour, suggestive of surface impurities.

Recently, further progress on surface composition has come from IR spectroscopy. Observations of Hyperion by Cruikshank and Brown (*Icarus* 50, 82; 1982) have revealed the presence of water ice bands, which differ in detail from those seen on Rhea. Measurements of the dark side of Iapetus by L. Lebofsky, M. Feierberg and A. Tokunaga (*Icarus* 49, 382; 1982) show an absorption feature thought to result from a hydrated silicate

100 years ago

POLLUTION OF THE ATMOSPHERE

THERE was a letter in *Nature* some time since, calling attention to the pollution of the atmosphere by the burning of coal; and it was calculated that in the year 1900, all animal life would cease, from the amount of carbonic dioxide; but the author had overlooked the fact that the rain is continually cleansing the atmosphere of this, and the fall of this rain on the ground, and the combination of this with various salts; besides the oceans alone would absorb their own bulk at normal pressure, but at an increased pressure of, say half a mile deep, would dissolve more than we are likely to need for hundreds of years.

But there are other products of combustion, or rather of incomplete combustion, that are not brought down in this manner by rain, as hydrogen and the hydrocarbons, chiefly marsh-gas and ethylene.

Since the year 1854 (as near as I can estimate) there has been burnt 10,000 million tons of coal; and if we say (in its consumption by household grates, leakage by gas-pipes, etc.) 1-100th escapes, then 100 million tons of hydrogen and hydrocarbons are floating in the

atmosphere or 1-10,000,000 part in bulk; if we say the average proportion of hydrogen to be .45, and of marsh gas .35, and of ethylene .4, we have .84 per cent of gases that are lighter than air, and it is more than probable that the law of diffusion of gases, as demonstrated with jars, does not apply to the atmosphere. The cases are not parallel: in the air we have unconfined space, pressure, and temperature diminishing infinitely, conditions favourable to the lighter and the gas with the greater amount of specific heat rising and maintaining its elevation, especially as we know that in large halls carbonic dioxide is found in larger quantities on the floor. According to Prof. Tyndall's researches, hydrogen, marsh gas, and ethylene have the property in a very high degree of absorbing and radiating heat, and so much so that a very small proportion, of only one thousandth part, had very great effect. From this we will have a marked influence on the climate of the world. The mountainous regions will be colder, the Arctic regions will be colder, the tropics will be warmer, and throughout the world the nights will be colder, and the days warmer. In the Temperate Zone winter will be colder, and generally differences will be greater, winds, storms, rainfall greater.

H. A. PHILLIPS

From *Nature* 27, 127, 7 December 1882.

Ian Gatley is at the UK Infrared Telescope Unit of the Royal Observatory, Edinburgh, in Hilo, Hawaii 96720.

REVIEW ARTICLE

Giant voids in the Universe

Ya. B. Zeldovich*, J. Einasto^{†‡} & S. F. Shandarin*

* Institute of Applied Mathematics, Moscow A-47, 125047, USSR

[†] Tartu Astrophysical Observatory, 202444 Estonia, USSR[‡] European Southern Observatory, 8046 Garching, FRG

Recent observations indicate that most galaxies are concentrated in superclusters consisting of galaxies, and clusters of galaxies, aligned along strings. Giant volumes exist between superclusters which are almost empty of visible objects. Theories of galaxy formation predict the formation of non-spherical superclusters and giant voids. Large-scale structure changes very slowly, so the currently observed structure reflects the whole history of galaxy formation and structural evolution.

OBSERVATIONS reveal that galaxies are distributed more or less randomly, showing a moderate tendency towards clustering which statistically can be expressed in terms of the galaxy correlation function¹. It has usually been assumed that field galaxies and clusters of galaxies are pretty well isolated from each other (island universes), and the correlation function was thought to reflect the size of clusters and the deviation from a purely random distribution.

Modern detectors have made it possible to determine the redshifts of many galaxies. The redshift, interpreted as the expansion speed of the Universe, is proportional to the distance of the galaxy from the observer. Combined with the apparent position in the sky, it is therefore possible to derive a three-dimensional distribution of galaxies in space.

The results of recent studies of the three-dimensional distribution of galaxies are surprising. Most (90%) galaxies are concentrated in strings, groups and clusters of galaxies²⁻⁶. Neighbouring strings form a connected network which can be called superclusters³. Superclusters are often flattened and fill only a small fraction of the total space (<10%), the remaining space between strings and superclusters contains almost no luminous matter^{3,6}. The diameter of voids between superclusters reaches $100 h^{-1}$ Mpc, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (refs 3, 7), and such voids appear to make up >90% of the total volume of the Universe.

This picture, particularly the presence of giant voids, differs completely from earlier expectations and raises several questions. First, is a revision needed of the fundamental cosmology, the Friedmann-Lemaître model of the Universe? Second, what does the observed structure tell us about the process of galaxy formation and about the very early Universe?

Observed random velocities of groups and clusters relative to the neighbouring string network seem to be small, of the order of 100 km s^{-1} (ref. 6). During the whole Hubble time a galaxy with such a small velocity could travel a distance <1 Mpc. Although the cluster network as a whole may have larger anomalous velocities^{1,8,9}, this does not alter the conclusion that the structure of the Universe changes slowly. Thus the present structure reflects the history of its formation and evolution.

We now review new observational data, and pay special attention to quantitative methods of analysing this data. Thereafter a scenario explaining the structure formation is outlined and competing scenarios are discussed.

All current theories of the formation of the structure are essentially gravitational clustering theories based on the theories of the expansion of the Universe¹⁰ and of density fluctuations¹¹. Both theories were first formulated in the

framework of general relativity. Their essentially newtonian content was understood much later¹².

Geometry of galaxy distribution

To a first approximation the clumpiness of galaxy distribution can be expressed in terms of a correlation function¹. The most popular two-point correlation function says, however, practically nothing about the characteristic pattern of galaxy distribution.

So far galaxy systems have been described only qualitatively. Our conviction that systems of galaxies of various types (groups, clusters, cluster chains) do exist is based on visual inspection of photographs and on galaxy counts. To build up quantitative theories of how the structure of the Universe formed, its structure must also be described quantitatively.

One possible way to do this is to use the so called 'cluster analysis' (see, for example, ref. 13). Draw a sphere (in the two-dimensional case a circle) of radius r around each dot representing galaxies. If the radius r is small enough, encircled regions are well separated from each other and the whole space is divided into isolated 'islands' containing galaxies and a continuous 'ocean' of empty space. When we take a larger radius, several neighbouring 'islands' join to form connected systems of galaxies. With an even larger radius, most galaxies join to form large connected systems. At a very large radius practically all galaxies join to form a single connected system.

The neighbourhood radius r at which galaxies form connected systems depend on the density of galaxies in space. With a small radius only the densest systems, cores of groups and clusters of galaxies, are picked up. Then, with increasing radius r neighbouring groups and clusters join to form elongated double or multiple clusters, thereafter still larger and less denser aggregates of galaxies will form connected systems.

The form and orientation of connected systems, their length, total volume and other properties depend on the pattern of the galaxy distribution.

Distribution of galaxies in space

Our analysis of the three-dimensional distribution of galaxies in space is based on a catalogue of all available redshifts of galaxies compiled by J. Huchra (Harvard-Smithsonian Center for Astrophysics). In the northern sky the catalogue is complete up to apparent magnitude 14.5, in the southern sky up to apparent magnitude 13.0. For galaxies brighter than the knee-point of the luminosity function (-21.0 absolute magnitude)

the catalogue is complete up to distances 130 and 65 Mpc (northern and southern sky, respectively). Throughout this review we use the Hubble constant $H = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Due to strong galactic absorption it is difficult or impossible to observe external galaxies in areas close to the Milky Way. Here, of course, data are not complete.

Several slices through the local supercluster are shown in Fig. 1, where supergalactic rectangular coordinates are also indicated and long galaxy strings between the Virgo and Coma clusters and regions void of galaxies can be seen.

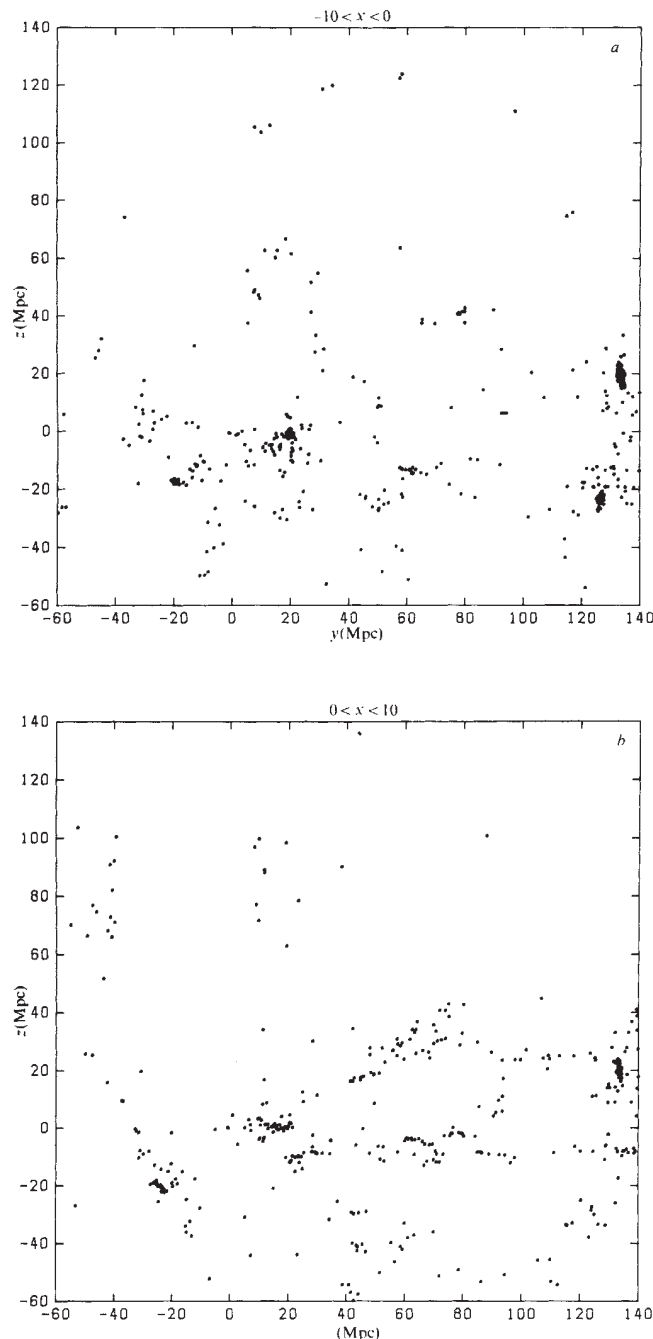


Fig. 1 Slices through the Local supercluster. Supergalactic rectangular coordinates are shown: the x and y axes lie in the supergalactic plane. The Galaxy is at the centre, the Virgo cluster lies in the y axis direction at $y = 20$ Mpc. Bright clusters of galaxies at $y = 130$ Mpc are the Coma cluster ($z = 20$ Mpc) and A1367 ($z = -10$ Mpc). Galaxies brighter than -19.5 absolute magnitude have been plotted.

Properties of systems of galaxies have been studied using the cluster analysis. For comparison, identical analysis was carried out for several model catalogues. One model was calculated following the method of Soneira and Peebles¹⁴. The second model was generated for a full random (Poisson) distribution of galaxies. The third model was taken from a N-body experiment, based on the adiabatic scenario of structure formation¹⁵. These model catalogues H, P, and A have approximately equal number and mean density of objects. The two-point correlation function was calculated for all model catalogues and for the observed (O) catalogue (the region centred at the Virgo cluster, cube side of 80 Mpc, limiting absolute magnitude -19.5 , total number of galaxies 866). The correlation function is plotted in Fig. 2. Functions are close to each other and only for the random P-distribution do we have zero correlation, as expected.

The maximal length of connected regions as a function of neighbourhood radius r is plotted in Fig. 3 for the observed and model distributions. At small r all curves corresponding to catalogues with normal correlation function have similar slopes. Connected regions are round or slightly elongated. Most regions found from the observational data correspond to known groups and clusters of galaxies.

At a certain critical radius (~ 1 Mpc) the slope of O and A curves abruptly increases. This corresponds to the transition from single to multiple clusters and cluster chains. With growing radius the length of chains rapidly increases and at some other critical radius (~ 5 Mpc in our case) it exceeds the size of the cube under study.

For the hierarchical H-catalogue the length increases slowly over the whole neighbourhood radius interval studied. For the P-catalogue the length increases at small r very slowly (the catalogue contains no built-in systems), but then the slope gradually increases.

Figure 4 shows the distribution of the number of galaxies in systems of different multiplicity for the observed and model catalogues. All histograms correspond to an neighbourhood radius $r = 5$ Mpc.

At this radius about half of all galaxies in the O and A catalogues form a single connected system. H and P catalogues have no large single connected systems. Galaxy systems, not connected to a single supersystem, have multiplicity < 64 . In the case of the O catalogue these galaxies are more or less evenly distributed among systems of different multiplicity. The H catalogue has too few singles and pairs (in comparison with the O catalogue). The distribution of small systems in the A and P catalogues is similar, peaking at singles.

Figures 3 and 4 can be interpreted as follows. Both random catalogues, H and P, behave completely differently from the observed O catalogue. Thus the observed distribution cannot be considered as a simple hierarchical one. In general, the A distribution has similar properties to the observed one. But we note two important differences.

First, in the A catalogue test particles that are not connected to the single large connected system, are distributed evenly like the particles in the P catalogue which can be considered as a field population. The O catalogue has no appreciable field population. Thus in regions of lower than the average matter density no galaxy formation takes place. The comparison of respective numbers of test particles suggests that about half of primordial matter can still be found in low density regions in some pre-galactic form.

The second difference lies in the fact that the A catalogue has no systems of intermediate size and mass. The observed catalogue has at this neighbourhood radius one large system (containing half the galaxies) and about 10 systems containing galaxies of multiplicity between 8 and 64 with length from 10 to 35 Mpc (mean value 20 Mpc). These systems are strings of galaxies and groups of galaxies, they contain some poor Zwicky¹⁶ clusters but no rich Abell¹⁷ clusters. The presence of systems of intermediate scale demonstrates that the Universe has some fine structure, which is lacking in the A model.

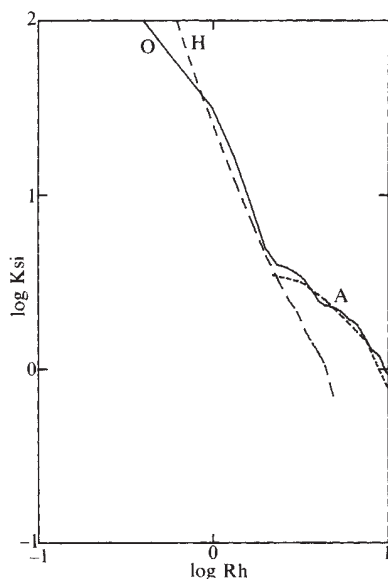


Fig. 2 Correlation function for the observed catalogue (O) around the Virgo cluster (cube of side 80 Mpc), for the catalogue generated by Soneira and Peebles method (H) and for the adiabatic model (A) (ref. 61).

The status of galaxies, not connected to systems of large or intermediate scale at this neighbourhood radius, is not clear. There are two possibilities, either these galaxies and small groups are outlying members of larger systems, or they constitute a true field population. To check these possibilities we have studied galaxies in a smaller and more complete region around the Virgo cluster. It is a cube of side 40 Mpc, has a limiting absolute magnitude of -16.5 , and contains 1,009 galaxies. We have calculated the distribution of galaxies among systems of various multiplicity for several values of the neighbourhood radii (1–5 Mpc). The results indicate that the relative number of singles and small systems quickly decreases with increasing radius, for $r = 3$ and 4 Mpc only 10 and 5% of all galaxies are located in systems of multiplicity < 8 , respectively. This test emphasizes that most small systems are outlying members of larger systems. The field population, if it exists, does not exceed 5–10% of the total number of galaxies.

Let us now estimate the volume occupied by systems of galaxies. Galaxy strings can be approximated by cylinders of length, found from the previous study of connected systems, and of radius, equal to the characteristic radius of galaxy groups and clusters. Nearby rich clusters of galaxies have following radii: Virgo, 2.2 Mpc; Perseus, 2.9 Mpc; Coma, 2.5 Mpc. The radius of groups found in our study of connected systems lies between 0.5 and 2 Mpc, the r.m.s. radius being 1.0 Mpc. We adopt as a mean outer radius of the cylinder $r = 2$ Mpc. The total length of galaxy strings in cubes around the Virgo cluster is 140 Mpc and 380 Mpc (smaller and larger cube respectively), so we found that within these cubes galaxy strings occupy 0.110 and 0.037 of the total volume.

To estimate the volume filled with galaxy systems in a larger region, only data on the distribution of bright galaxies (absolute magnitude < -21) can be used. For this we have divided the region under study into cubic cells of various size. We note that at cell size 5 Mpc relative volume of cells occupied by bright galaxies is approximately equal to the relative volume occupied by galaxy systems. Using this method we have found that the filling factor of a cube of side 200 Mpc centred on the Virgo cluster is only 0.01. Thus the density of galaxies outside superclusters should be very small.

Where are the low density areas? Figure 1 shows that at high supergalactic z coordinates the galaxy density is much lower than near the supergalactic plane $z = 0$. The almost complete

absence of galaxies in front of the Hercules supercluster is well known⁵. To study the galaxy distribution in vertical supergalactic direction we have found the number of bright galaxies in rectangular sheets of sizes $200 \times 200 \times 20$ Mpc. One sheet is centred at the Virgo cluster, others lie upwards and downwards. The distribution of numbers of galaxies in these sheets as a function of z is given in Fig. 5. We see that the number of galaxies rapidly decreases with increasing distance from the supergalactic plane. A similar distribution in the x and y directions shows no systematic trend.

The rapid decrease of the number of galaxies in the z direction is partly due to the galactic absorption: the zone of avoidance crosses the sky near supergalactic poles. A rough estimate shows that the number of galaxies at extreme z values may be underestimated by a factor of three. The corrected distribution is also plotted in Fig. 5, as well as the distribution of rich clusters of galaxies having redshift $< 7,000 \text{ km s}^{-1}$. Nearby clusters of galaxies are very strongly concentrated towards the supergalactic plane and belong to three well studied superclusters: the Local, Coma, and Perseus, as well as to several other previously unrecognized superclusters.

This sheet of superclusters can be considered as a wall shared by two local cells of the Universe, the northern and the southern¹⁸. The Hercules supercluster, the supercluster around Abell cluster A539 and several southern superclusters form side walls of these cells. The far side of the local cells can be located only approximately, due to the lack of observations near the galactic plane. Both local cells have a diameter of ~ 200 Mpc.

Available data suggest that the interior of local cells contains no rich clusters of galaxies. In this respect they are similar to cells found by Jõeveer *et al.*² behind the Perseus supercluster and by Kirschner *et al.*⁷. However, there are also some poor Zwicky¹⁶ clusters here¹⁹. Our study confirms the results obtained by Balzano and Weedman²⁰ indicating that Kirshner's void is not completely empty. Zwicky clusters as well as galaxies located in cell interiors are not randomly spaced. As in superclusters, galaxies and clusters form long strings here. The basic differences between galaxy strings in superclusters and in cell interiors lie in the population and the spacing: in cell interiors

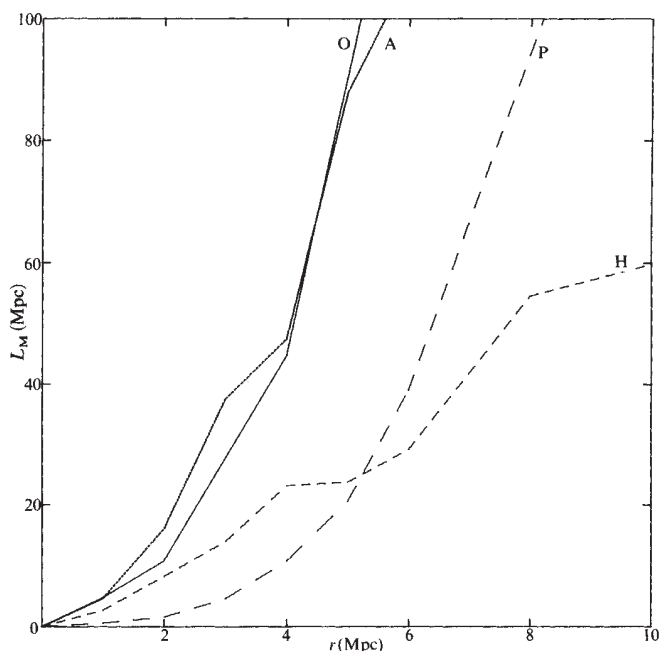


Fig. 3 Maximal length L_M of connected regions as a function of neighbourhood radius r for four catalogues⁶¹.

the string population is smaller and the spacing between neighbouring strings is larger.

From available data the luminosity density in different structures can be calculated. As luminosity indicators we use bright galaxies. The total luminosity in solar units for standard luminosity functions is

$$L = n^* \times 3.6 \times 10^{11} L_{\odot}$$

where n^* is the number of galaxies brighter than $M = -21.0$ (ref. 2). We found that for the average luminosity density $5 \times 10^7 L_{\odot} \text{Mpc}^{-3}$, the mean density in cell interiors is about 3 times lower, but in galaxy strings about 70 times higher. An upper limit for the luminosity density outside connected regions is 10 times lower than the average luminosity density. Thus the luminosity density contrast between galaxy strings and 'empty' space is at least three orders of magnitudes.

Cluster analysis gives us the possibility of picking up and studying individual connected systems. This study demonstrates that most systems are elongated and look like cluster strings in the Perseus supercluster^{2,3}. In some cases, however, small sheets of galaxies have been found; one such sheet is connected with the Virgo cluster. Three orthogonal views of this system are given in Fig. 6. The interpretation of sheets of this type is still open. No large sheets of galaxies have been found so far.

Compatibility of the observed structure with the Friedmann-Lemaître model

Large voids between superclusters have a volume of $\sim 10^6 \text{Mpc}^3$ (refs 3, 7). In an expanding Universe the diameter of voids expressed in redshift units remains approximately constant (exactly constant in the case of a non-decelerating Universe), thus we can adopt as a first approximation $\Delta z = 0.05$. Let us assume that the cellular structure was formed at $z = 3-10$, as indicated by a comparison of densities in various structural elements and by the cutoff of quasar redshifts at $z = 3.5$ (refs 3, 21). If this is the case, then within the horizon our Universe should contain $\sim 10^6$ cells.

There is little or no evidence for the presence of still larger structural units in the Universe. Cell dimensions in comparison with the Universe as a whole are still small, thus we see no need for a revision of the fundamental principles of cosmology. Our confidence in the classical model is strengthened by small temperature fluctuations of the microwave radiation, indicating small density fluctuations at the epoch of decoupling of radiation and matter ($z \approx 1,300$) (refs 22, 23).

How the structure of the Universe formed

At present it is difficult to infer the type, the spectrum and the amplitude of initial perturbations from general physical principles. So theorists consider all possibilities. The most developed scenarios of the formation of the large-scale structure are based on two different assumptions about the character of perturbations in the pre-decoupling era²⁴.

First, at an early stage both matter and radiation density were perturbed so that their ratio was constant in space and time. Because this ratio is conserved in adiabatic processes such perturbations are termed adiabatic.

Second, it is supposed that the matter density was perturbed but the radiation density was not. These perturbations are known as isothermal since in this case temperature is assumed to be the same everywhere.

Both scenarios have important features in common: the initial perturbations on scales larger than galaxy ones are small, the large density contrast and the large-scale structure develops due to the gravitational forces. In an expanding Universe small density fluctuations can grow only if gravitation amplifies them. In this sense both the adiabatic and the isothermal scenarios are gravitational clustering theories, however, the principal

differences between them lie in the scale of post-decoupling perturbations.

Adiabatic scenario

Adiabatic fluctuations on scales smaller than a characteristic mass $M_c \approx 10^{13}-10^{15} M_{\odot}$ were damped before decoupling due to photon viscosity²⁵; after decoupling one finds practically no fluctuations on scales smaller than this limit. Perturbations of the scale M_c had the largest amplitude, thus under- and over-densities of this scale developed first. Galaxies and clusters of galaxies formed in dense regions after the gas collapsed and cooled. The formation of dense regions and galaxies started at a relatively late epoch $z = 3-10$.

The principal feature of the adiabatic scenario lies in the fact that the formation of large-scale units occurs in a cold environment. The gas pressure is negligible until the final stages of formation of large-scale structures. Numerically this fact is expressed in terms of the ratio of the Jeans mass of the neutral gas after decoupling, $M_J \approx 10^6 M_{\odot}$, to the characteristic mass of adiabatic perturbations, $M_c \approx 10^{13}-10^{15} M_{\odot}$: $M_J/M_c \approx 10^{-7}-10^{-9}$. In the first approximation we can neglect the temperature and gas pressure, assuming $M_J/M_c = 0$.

The evolution of perturbations in a pressureless environment is now well established. Before giving more details on the formation of the large-scale structure let us outline the main processes leading to galaxy formation in the adiabatic scenario^{24,26,27}.

Very flattened clouds—'pancakes'—with masses about $M_c \approx 10^{13}-10^{15} M_{\odot}$ were formed first²⁸. At its arrival to a pancake the gas is heated and compressed in the shock wave, which represents the boundary of the pancake. Inner parts of pancakes cool quickly due to radiative losses of their thermal energy causing an increase in density²⁹. So in inner parts of pancakes the conditions become favourable for galaxy formation^{26,27}. Shock waves on the boundaries of pancakes favour the generation of rotation, thus spiral galaxies could form easily³⁰. The general trend in the evolution on galaxy scales after pancake

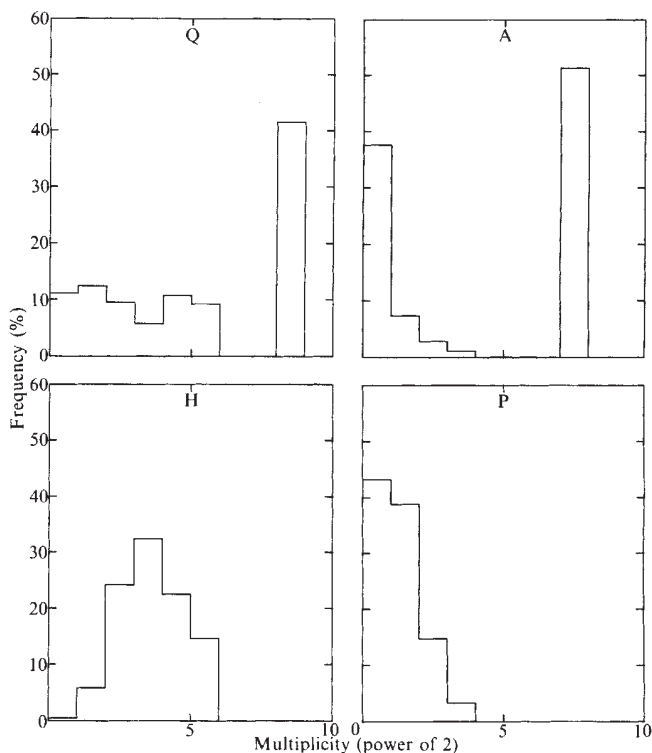


Fig. 4 Distribution of galaxies according to the multiplicity of the system. Neighbourhood radius $r = 5 \text{Mpc}$. Multiplicity is expressed in powers of 2 (5 corresponds to $2^5 = 32$ and so on)⁶¹.

formation is the fragmentation of pancakes into smaller units. Thus the adiabatic theory is often called the fragmentation scenario.

Large-scale structure and the catastrophe theory

We now consider the formation of larger-scale objects. We believe that the processes of galaxy formation do not have a major influence on the motion of matter on scales of rich clusters and superclusters. As noted earlier, the growth of adiabatic perturbations has one important property: the evolution proceeds at low temperature of self-gravitating medium. This evolution has a simple approximate analytical solution²⁸. As the smooth initial perturbations develop into the nonlinear stage, dense regions will be built up by the concentration of matter into caustics—the intersection of particle trajectories.

The formation of caustics in a swarm of dust particles has been demonstrated by Lifshitz and Khalatnikov³¹ and later studied by Grishtshuk³². Initially it was thought that this case is physically not realistic because an arbitrarily small pressure removes sharp caustics. However, the idealized case is a very good approximation to the actual situation. The geometry of dense regions is closely related to that of caustics in a pressureless medium.

Geometrical properties of caustics can be studied mathematically using the catastrophe theory^{33,34}. Because adiabatic perturbations are potential ($\text{rot } \vec{v} = 0$), the study of singularities in the potential flows is of great interest.

As shown by Arnold and co-workers³⁵ following an early suggestion by Zeldovich²⁸, pancake-like singularities known in the catastrophe theory as lips formed first. At a later epoch pancakes grew, intersected and formed a fairly pronounced cellular structure³⁶. Matter continued to flow towards and within pancakes, so that some time later higher-order singularities, lines and knots, formed. All this happened in a similar way both in gaseous medium and in some collisionless cold (neutrino) medium. This phase ended with a cellular structure of high density sheets, strings in sheets and knots connected by strings. At this stage dense regions fill <10% of the total volume of the Universe, while isolated low density regions (cell interiors) fill the remaining major part of the volume of the Universe.

Numerical experiments demonstrate^{15,37} that the density is highest in knots (zero-dimensional structures), high in strings (one-dimensional structures) but much lower in sheets (two-dimensional structures). Subsequent evolution of the structure depends on three processes, the formation of galaxies, the gravitational redistribution (clustering) of galaxies, and the expansion of the Universe.

One comment is necessary here: galaxies are easier born in dense and cold gas. Thus it is natural to expect that no galaxies will be formed in cell interiors and that an intensive galaxy formation would occur along the lines and in the knots. It is not evident whether the density in all sheets is sufficient for galaxy formation or not. Perhaps some gaseous sheets remain in gaseous form and can be seen as absorption lines in quasar spectra^{38,39}.

Later gravitational clustering tries to destroy the cellular structure, the rate of which depends on the mean density of the Universe. In a low density Universe cellular structure is maintained for a long time, in a closed Universe this structure disappears in a time scale comparable with the Hubble time, thereafter galaxies clump to clusters of very high mass^{15,37}.

Neutrino dominated Universe

Perhaps the weakest point in the adiabatic scenario is its need for too large an amplitude of density perturbations at the decoupling era: $\delta\rho/\rho \approx 10^{-3}$ if $\Omega = 1$ and $\delta\rho/\rho \approx 10^{-1}$ if $\Omega = 0.02$ (ref. 40). As noted already by Silk²³, density fluctuations at the epoch of decoupling correspond to similar angular fluctuations of the temperature of the microwave background, $\delta T/T \sim$

$1/3\delta\rho/\rho$. On the other hand, observations give an upper limit of temperature fluctuations of the order 10^{-4} (refs 22, 23).

In an open Universe density fluctuations grow extremely slowly. To obtain the observed density contrast, it is also necessary to have at the epoch of decoupling high-density perturbations in contrast to small temperature fluctuations of the microwave background.

One can infer from these arguments that $\Omega = 1$ is necessary for the adiabatic scenario. However, the data on helium and deuterium abundances strongly support the model of a low baryon density. Deuterium cannot be produced in stars like heavy elements, the whole deuterium observed now must have originated during the primordial fireball stage of the Universe. Deuterium production rate is extremely sensitive to the density of baryons in the Universe. Observed deuterium abundance suggests the baryon density 0.01–0.1 of the closure density⁴¹, that is the Universe should be open by a large margin.

Last but not least, flat rotation curves of galaxies, the dynamics of double galaxies and of clusters of galaxies indicates that giant galaxies and clusters of galaxies are embedded in huge invisible coronas^{42,43}. The mass of the coronas exceeds the mass of visible galactic populations by a factor of 10 or more.

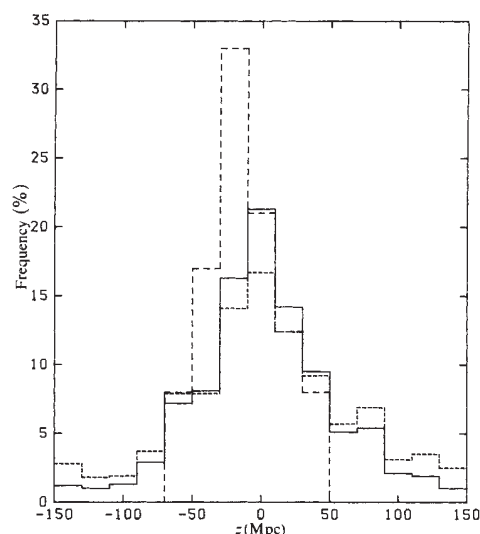


Fig. 5 Distribution of galaxies and rich clusters of galaxies in supagalactic z coordinate. Solid line, uncorrected galaxy distribution; dotted line, galaxy data corrected for incompleteness; dashed line, the cluster distribution.

This controversy would be solved if the Universe were neutrino dominated with the neutrino mass $m \approx 10$ eV. Neutrino gas does not interact with radiation, thus perturbations in the neutrino gas could develop much earlier than in the baryon dominated Universe and could have the necessary amplitude. Baryon gas is bound to radiation and has smaller density fluctuations, after decoupling it simply flows to gravitational wells formed in the neutrino gas.

Thus in the neutrino dominated Universe one has low baryon density $\Omega_b \approx 0.01$ –0.1 while the total density is close to the closure once $\Omega_r \approx \Omega_b \approx 1$.

The formation of the structure in a neutrino dominated Universe is, essentially, an adiabatic scenario^{44–51}. The initial ratio of baryons to neutrinos is the same everywhere (the entropy is constant), small-scale fluctuations are damped, the characteristic mass of objects to form first is $10^{15} M_\odot$ as in the conventional adiabatic scenario.

However, the structure of pancakes is more complicated: inside a neutrino pancake there is a pancake made of baryons which is similar to that described earlier.

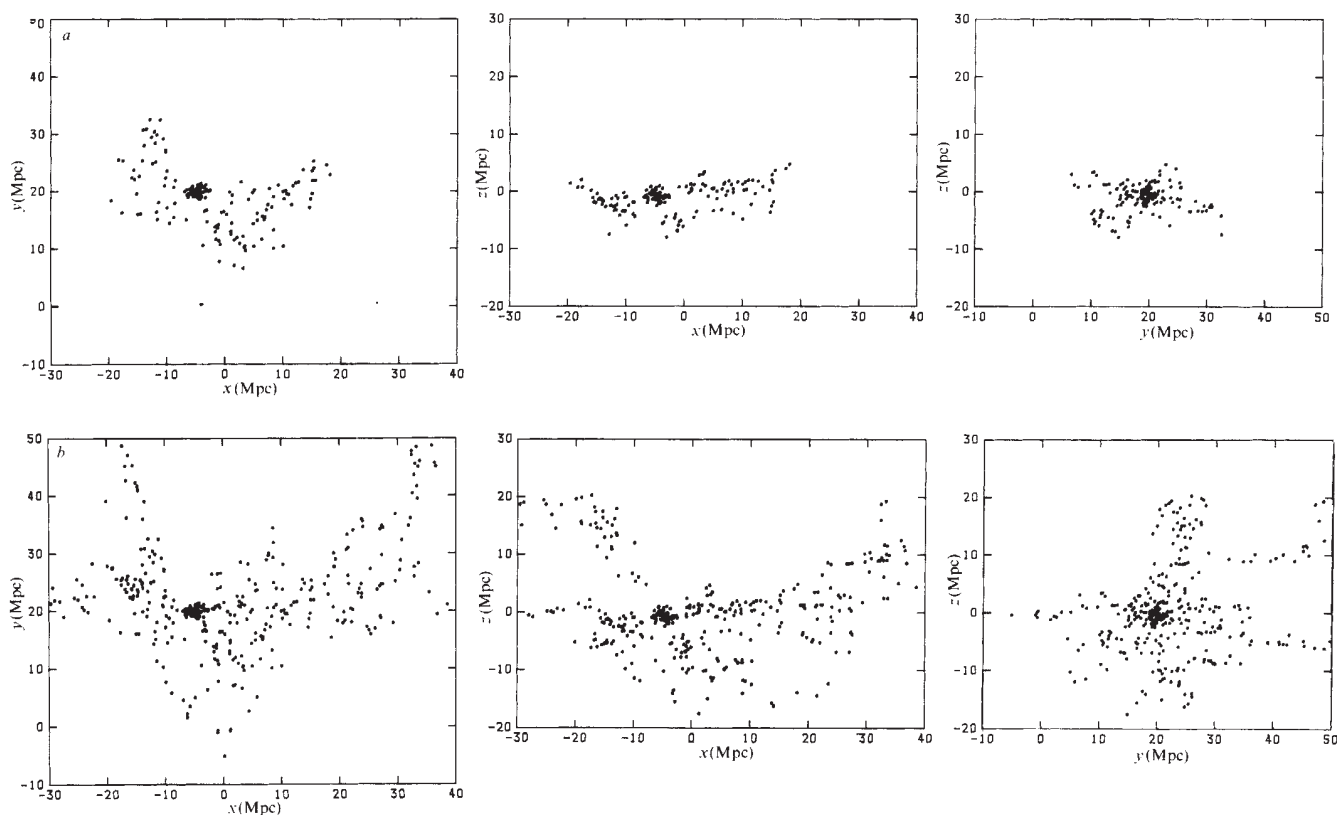


Fig. 6 Three views of a connected system around the Virgo cluster⁶¹: *a* corresponds to the neighbourhood radius $r = 4$ Mpc; *b* to the radius $r = 5$ Mpc. *a* shows two sheets of galaxies in both sides of the Virgo cluster, *b* demonstrates the presence of a string network around the cluster.

Isothermal scenario

According to the isothermal scenario, small-scale density fluctuations were not damped. Thus, it is natural to think that the mass of the first objects to be formed was limited by the gas pressure—by the respective Jeans mass after decoupling 10^5 – $10^6 M_\odot$ (refs 49, 52). Thus globular cluster-sized objects were the first primordial building units formed soon after decoupling, $z \approx 10^3$. During the subsequent evolution larger building blocks, galaxies and clusters of galaxies, were built up step-by-step by gravitational clustering.

The formation of the first bound objects of mass $10^6 M_\odot$ takes place in the environment when one cannot neglect the influence of pressure. Therefore they must have a shape that is almost spherical. Part of gravitational energy liberated during their formation transforms into chaotic motion, having the role of 'temperature' of the 'gas' of these primordial objects. Hence, according to the isothermal scenario, the formation of large-scale structural units proceeds in a high-temperature 'gaseous' environment. The 'gas' pressure was active during the whole process of the formation of the structure so that it is difficult to build up strongly anisotropic structures.

Extensive series^{53–55} of three-dimensional N-body calculations in the framework of the isothermal model reproduced the observational correlation function. The general appearance of particle distribution was claimed to be similar to the observational distribution of galaxies. However, it seems that in details the calculated distribution is more amorphous than the observational one.

Tepid and other cosmological models

Another scenario of the formation of the large-scale structure is suggested by Ostriker and Cowie⁵⁶. They argued that the formation of large-scale second generation objects (superclusters around voids) may be due to the nuclear energy released during the explosions of first generation objects. These

explosions can give rise to expanding shock wave shells which ignite star and galaxy formation in the remaining gas.

Earlier it was recognized that total energy released during explosions of primordial objects is comparable with the energy contained in the microwave background radiation. This point has led Rees (refs 57, 58 and references therein) to the hypothesis that microwave radiation can be of nuclear origin and thus that there is no need for a hot Big Bang.

Both scenarios predict many dead remnants of first generation stars, therefore, they easily solve the problem of missing mass.

Observable differences expected in different formation scenarios

In the isothermal case dominant post-decoupling perturbations had a scale corresponding to a mass of $10^6 M_\odot$. Star cluster like objects formed very quickly after decoupling and most of the primordial gas went into stars at an early stage. Large-scale structures, such as superclusters and large voids, were due to large-scale perturbations, which needed more time to develop and formed gradually. Galaxies should have a tendency towards clustering, but could be found everywhere between clusters and strings. Strings should be short and not connected with other structural elements, voids should have moderate sizes. These expectations have been confirmed by numerical simulations^{53,54}.

In the adiabatic case small-scale perturbations were damped. The Universe evolved slowly and was gaseous for a long time. In contrast to the isothermal scenario, clusters and groups of galaxies formed at a later epoch along the initially continuous pre-existing network of gaseous strings and knots. The density of gas outside this network was small, thus no galaxies should have been formed there^{36,59}. Some ionized gas and dark matter is still in 'empty' areas, because it is difficult to evacuate these large regions completely⁶⁰. This gas should have had primordial abundance.

In the Ostriker and Cowie⁵⁶ version one expects to have rather empty cell interiors, also filled with some ionized gas that should have a normal chemical composition.

The expected differences are of two types: (1) the composition of the gas outside superclusters; and (2) the fine structure of superclusters and 'empty' areas.

Gas layers in cell walls can be studied using absorption line spectra of distant quasars as suggested by Oort³⁸ (see also ref. 39). Here not only the presence of gas but also its chemical composition is important. Geometrical properties of alignment of galaxies and clusters of galaxies can be best studied in the vicinity of our own Galaxy.

Problems

The history of the formation of large-scale structure critically depends on basic cosmological constants, the time scale available (Hubble constant H), the overall density of matter (the ratio of density to the closure density $\Omega = \rho/\rho_c$), and last but not least, on the nature of the dark matter (is it baryon matter or neutrino gas?).

A very important observational constraint is the temperature fluctuation of the microwave background radiation. Here not only the upper limit but also the actual fluctuations and their angular size are needed.

Conventionally observational studies in cosmology are based on data of objects located at high 'cosmological' redshifts. The study of superclusters in our vicinity has shown that important cosmological information can be received also by studying the distribution of nearby galaxies. Here we can study the distribution in greater detail.

To build up a quantitative theory of the formation of the structure in the Universe, the structure itself must be characterized quantitatively. Some weak points of principal structure formation scenarios that are already known are now considered.

The adiabatic scenario predicts the formation of structures of three types: surfaces, lines and dots. Numerical experiments suggest that the density is highest in dots, relatively high in lines, much lower in surfaces and very low in regions outside these singularities. Were galaxies formed in all of dense regions? If so, should we expect to observe all three kinds of galaxy system: pancakes, strings and clusters? The tests carried out so far demonstrate the presence of strings and clusters, the evidence for pancakes is controversial. Does this indicate that galaxy formation occurs basically at the structures having highest densities—in lines and dots?

The adiabatic scenario also predicts the formation of a simple cellular structure³⁶. Observations indicate that the real distribution is much more complex. Rich clusters of galaxies (of the type included in the Abell catalogue) and connecting cluster chains surround regions, which are void of rich cluster. Voids of this type have a diameter of 200 Mpc. These voids are not completely empty, they are intertwined by a lattice of galaxy strings. Near the rich cluster the network of galaxy strings is much more closely spaced. How can we explain this hierarchical structure? Why do the properties of the fine structure (small galaxy strings) depend on the large-scale environment (proximity to rich clusters)?

The severest problem encountered by the isothermal scenario is the amorphous character of the calculated structure. This conclusion is based on the visual inspection of the results of N-body calculations. To draw more quantitative conclusions, one must await objective checks.

Another weak point in the isothermal scenario is the galaxy formation. As Silk has pointed out⁴⁶, there are several arguments (such as high central concentration demanding dissipation of extra energy, and radial gradient of chemical composition favouring *in situ* formation of heavy elements) demanding that galaxies have been formed directly from gas, not by clustering of ready stellar systems of lower mass.

Astronomers usually question which formation scenario is right. But both basic scenarios have weak points—perhaps it would be more correct to ask, which physical processes have led to the formation of structure in the Universe?

To answer this question, more observational data are needed. In particular, the study of 'empty' regions is important. So far astronomers have concentrated on rich clusters and other prominent features. The study of empty regions requires the coverage of large areas in the sky up to a fairly faint apparent magnitude, and thus much observing time. The first line of attack should be the study of the central regions of local cells. This can be done using telescopes of moderate size (of the 1.5–2 m class).

We thank many colleagues, in particular Professors V. I. Arnold, R. H. Miller, J. H. Oort and M. J. Rees, and Drs A. G. Doroshkevich, J. E. Hartley, E. Saar and R. A. Sunyaev for helpful discussions, also Drs J. Huchra and H. Rood for unpublished redshift compilations, and Professor R. H. Miller and Dr A. Klypin for permission to use preprint material. J. E. benefited from the hospitality of the European Southern Observatory, where a considerable part of this paper was written.

1. Peebles, P. J. E. *The Large-Scale Structure of the Universe* (Princeton University Press, 1980).
2. Joeveer, M., Einasto, J. & Tago, E. *Mon. Not. R. astr. Soc.* **185**, 357 (1978).
3. Einasto, J., Joeveer, M. & Saar, E. *Mon. Not. R. astr. Soc.* **193**, 353 (1980).
4. Gregory, S. A., Thompson, L. A. & Tift, W. G. *Astrophys. J.* **243**, 411 (1981).
5. Tarenghi, M., Tift, W. G., Chincarini, G., Rood, H. J. & Thompson, L. A. *Astrophys. J.* **234**, 793 (1979).
6. Tully, R. B. *Astrophys. J.* **257**, 389 (1982).
7. Kirshner, R. P., Oemler, A. Jr & Shectman, S. A. *Astrophys. J. Lett.* **248**, L57 (1981).
8. Doroshkevich, A. G. & Shandarin, S. F. *Astr. Zh.* **54**, 734 (1977).
9. Zeldovich, Ya. B. & Sunyaev, R. A. *Rev. Astr. Astrophys.* **18**, 537 (1980).
10. Friedmann, A. Z. *Physik* **10**, 377 (1922).
11. Lifshitz, E. M. *Zh. eksp. teor. Phys.* **16**, 587 (1946).
12. Zeldovich, Ya. B. *A. Rev. Fluid. Mech.* **9**, 215 (1977).
13. van Ryzin, J. (ed.) *Classification and Clustering* (Academic, New York, 1977).
14. Soneira & Peebles, P. J. E. *Astr. J.* **83**, 845 (1978).
15. Klypin, A. A. & Shandarin, S. F. Preprint N131 (Institute of Applied Mathematics, USSR Academy of Sciences, 1981).
16. Zwicky, F., Wild, P., Herzog, E., Karpowicz, M. & Koval, C. *Catalogue of Galaxies and Clusters of Galaxies, 1–6* (California Institute of Technology, Pasadena, 1961–68).
17. Abell, G. *Astrophys. J. Suppl.* **3**, 211 (1958).
18. Einasto, J., Corwin, H., Huchra, J., Miller, R. H. & Tarenghi, M. Preprint (Tartu Astrophysical Observatory, 1982).
19. Tago, E., Einasto, J. & Saar, E. Preprint (Tartu Astrophysical Observatory, 1982).
20. Balzano, V. A. & Weedman, W. *Astrophys. J. Lett.* **255**, L1 (1982).
21. Osmer, P. *Astrophys. J.* **253**, 28 (1982).
22. Patridge, R. B. *Phys. Ser.* **21**, 614 (1980).
23. Silk, J. *Proc. 10th Texas Symp. Rel. Astrophys.*, 188 (1981).
24. Zeldovich, Ya. B. & Novikov, I. D. *Structure and Evolution of the Universe* (Chicago University Press, in the press).
25. Silk, J. *Astrophys. J.* **151**, 459 (1968).
26. Zeldovich, Ya. B. *The Large-Scale Structure of the Universe* (eds Longair, M. S. & Einasto, J.) 399 (Reidel, Dordrecht, 1978).
27. Doroshkevich, A. G., Shandarin, S. F. & Saar, E. *Mon. Not. R. astr. Soc.* **184**, 643 (1978).
28. Zeldovich, Ya. B. *Astr. Astrophys.* **5**, 84 (1970).
29. Doroshkevich, A. G., Zeldovich, Ya. B. & Novikov, I. D. *Soviet Astr.* **11**, 233 (1967).

30. Doroshkevich, A. G. *Astrophys. Lett.* **14**, 11 (1973).
31. Lifshitz, E. M. & Khalatnikov, U. M. *Adv. Phys.* **12**, 185 (1963).
32. Grishchuk, L. P. *Zh. eksp. teor. Fiz.* **53**, 1699 (1967).
33. Thom, R. *Structural Stability and Morphogenesis* (Benjamin, New York, 1975).
34. Arnold, V. I. *Comm. Pure. Appl. Math.* **39**, 557 (1976).
35. Arnold, V. I., Shandarin, S. F. & Zeldovich, Ya. B. *Geophys. Astrophys., Fluid Dyn.* **20**, 111 (1982).
36. Zeldovich, Ya. B. *Pisma Astr. Zh.* **195** (1982).
37. Doroshkevich, A. G. et al. *Mon. Not. R. astr. Soc.* **192**, 321 (1980).
38. Oort, J. H. *Astr. Astrophys.* **94**, 359 (1981).
39. Doroshkevich, A. G. & Shandarin, S. F., *Mon. Not. R. astr. Soc.* **179**, 95 (1977).
40. Rozgacheva, F. K. & Sunyaev, R. A. *Pisma Astr. Zh.* **7**, 323 (1981).
41. Wagoner, R. V. *Astrophys. J.* **179**, 343 (1973).
42. Einasto, J., Kaasik, A. & Saar, E. *Nature* **250**, 309 (1974).
43. Gallagher, J. E. & Faber, S. M. *A. Rev. Astr. Astrophys.* **17**, 135 (1979).
44. Doroshkevich, A. G., Khlopov, M. Yu., Sunyaev, R. A., Szalay, A. S. & Zeldovich, Ya. B. *Proc. 10th Texas Symp. Relativistic Astrophysics* (1981).
45. Sato, H. *Proc. 10th Texas Symp. on Relativistic Astrophysics*, 43 (1981).
46. Silk, J. *Cosmology and Fundamental Physics* (Vatican Study Week, 1981).
47. Bisnovaty-Kogan, G. S. & Novikov, I. D. *Astr. Zh.* **57**, 899 (1980).
48. Doroshkevich, A. G. & Khlopov, M. Yu. *Astr. Zh.* **58**, 913 (1981).
49. Bond, J. R., Efstathiou, G. & Silk, J. *Phys. Rev. Lett.* **45**, 1980 (1981).
50. Sunyaev, R. A. & Zeldovich, Ya. B. *Astr. Astrophys.* **20**, 189 (1981).
51. Zeldovich, Ya. B. & Shandarin, S. F. *Pisma Astr. Zh.* **8**, 259 (1982).
52. Dicke, R. H. & Peebles, P. J. E. *Astrophys. J.* **194**, 838 (1968).
53. Aarseth, S. J., Gott, J. R. & Turner, E. L. *Astrophys. J.* **228**, 644 (1979).
54. Efstathiou, G. & Eastwood, G. W. *Mon. Not. R. astr. Soc.* **194**, 504 (1981).
55. Miller, R. H. *Astrophys. J.* (in the press).
56. Ostriker, J. P. & Cowie, L. L. *Astrophys. J. Lett.* **243**, L127 (1981).
57. Rees, M. *Nature* **275**, 35 (1978).
58. Rees, M. *Observational Cosmology* (eds Gunn, I. E., Longair, M. S. & Rees, M. J.) 259 (Geneva Observatory Publication, 1978).
59. Zeldovich, Ya. B. & Shandarin, S. F. *Pisma Astr. Zh.* **8**, 131 (1982).
60. Einasto, J., Joeveer, M. & Saar, E. Preprint (Tartu Astrophysical Observatory, 1979).
61. Einasto, J., Klypin, A. A., Saar, E. & Shandarin, S. F. Preprint (Tartu Astrophysical Observatory, 1982).

ARTICLES

Absolute age of formation of chondrites studied by the ^{87}Rb – ^{87}Sr method

J-F. Minster, J-L. Birck & C. J. Allègre

Laboratoire de Géochimie et de Cosmochimie, Département des Sciences de la Terre, Institut de Physique du Globe de Paris, Universités Pierre et Marie Curie et Paris VII, 4, Place Jussieu-75230 Paris Cedex 05, France

The ^{87}Rb – ^{87}Sr isochrons for H-, E- and LL-, chondrites are undistinguishable from each other. The joint isochron yields an age of $4,498 \pm 15$ Myr if the rubidium decay constant $\lambda^{87}\text{Rb} = 1.42 \times 10^{-11} \text{ yr}^{-1}$, and a strontium initial ratio $^{87}\text{Sr}/^{86}\text{Sr} = 0.69885 \pm 0.00010$. For the age to agree with U–Th–Pb data, $\lambda^{87}\text{Rb}$ would have to be changed to $1.402 \pm 0.008 \times 10^{-11} \text{ yr}^{-1}$.

DATING of meteorites helps to decipher the mode of formation and the history of these extraterrestrial objects. This also provides constraints—mainly as a starting point—for the evolution of the terrestrial planets. In this respect, chondrites are particularly useful because they are chemically undifferentiated relative to the solar abundances and have had a fairly simple history. Thus it is thought that their age of formation corresponds to that of the Solar System. An absolute value for this age can in principle be obtained from the radioactive chronometers and we aim here to give a 'best' ^{87}Rb – ^{87}Sr age of chondrites and discuss its interpretation.

In this method, the $^{87}\text{Sr}/^{86}\text{Sr}$ isotopic ratio of a series of objects is plotted against their $^{87}\text{Rb}/^{86}\text{Sr}$ atomic ratio. If these objects have been formed simultaneously with the same $^{87}\text{Sr}/^{86}\text{Sr}$ initial isotopic composition, and if they evolved since as closed systems (without external exchange of Rb and Sr), they should define a straight line in such a diagram. This line is called an isochron; its slope yields the age of the objects and its intercept gives the strontium initial ratio (Fig. 1). Besides the uncertainties introduced by the analytical determination of $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{87}\text{Rb}/^{86}\text{Sr}$, the only source of error is that of the decay constant of the parent nuclide ^{87}Rb . Because its decay energy is very low, its direct determination is relatively unprecise. The current value is $\lambda^{87}\text{Rb} = 1.42 \pm 0.01 \times 10^{-11} \text{ yr}^{-1}$ (refs 1, 2). Another way to obtain this constant is to compare ages obtained by the ^{87}Rb – ^{87}Sr method with those obtained by the U–Pb method^{1,3}. Indeed, as the decay constant of ^{235}U is much larger than that of ^{87}Rb , it is easier to measure and its uncertainty is $<0.3\%$ (ref. 4). We now compare ^{87}Rb – ^{87}Sr and U–Th–Pb ages for chondrites and indicate that, in order for the ages to agree, the ^{87}Rb decay constant would have to be changed.

Compared with dating terrestrial or lunar samples, dating of meteorites presents a major difficulty: because their origin is uncertain, it is impossible to know in advance whether they are genetically related. The best approach is therefore to analyse a set of meteorites and, if an isochron is obtained, to infer that they are genetically related, or at least formed within a short interval of time compared with the uncertainty of the dating. In the early works on ^{87}Rb – ^{87}Sr dating of meteorites, objects of widely varying compositions, including chondrites and basaltic achondrites, were analysed^{5–9}. As the latter are differentiated objects, largely deficient in Rb compared with chondrites, fairly well-defined 'isochrons' were obtained which gave an age of 4,560 Myr, with an error limit of the order of 150 Myr.

Clearly, however, such drastically different objects cannot be genetically related in a simple way. A 'primitive' ^{87}Rb – ^{87}Sr isochron should use only chondrites. Moreover, chondrites can be classified into chemical groups¹⁰. Oxygen isotopes studies have recently provided some evidence in favour of this classification: indeed, the chondrites from a given group tend to have similar oxygen isotopic composition¹¹. Though this

result is not systematic, it indicates that the chondrites from a given group originated from a single reservoir of the solar nebula. It is thus preferable to date chondrites group by group. The first attempt of such an analysis¹² has been followed by a systematic study by Wetherill and his co-workers^{13,14}. This resulted in a set of isochrons for carbonaceous, ordinary and enstatite chondrites which were grouped by Wetherill¹⁵ into a

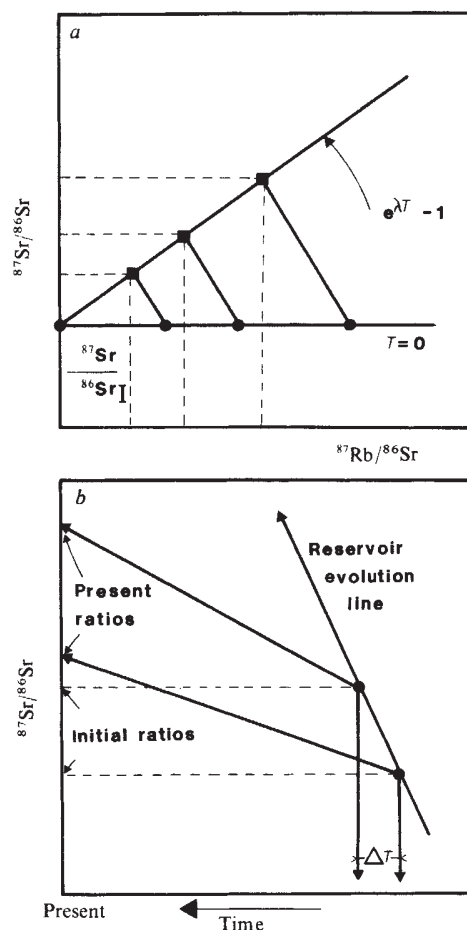


Fig. 1 *a*, Schematic evolution of the $^{87}\text{Sr}/^{86}\text{Sr}$ isotopic compositions of a series of objects formed simultaneously with different Rb/Sr ratios. Such objects plot on a straight line called isochron whose slope yields the 'age' of the objects. *b*, Evolution with time of the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio in a reservoir. The higher the Rb/Sr ratio in the reservoir, the faster this evolution. The difference between the initial ratio of two objects formed from a single reservoir can be used to calculate an age difference.

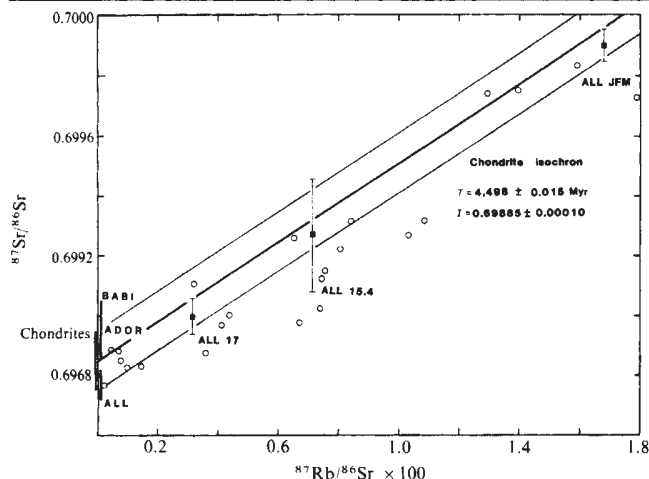


Fig. 2 Plots of $^{87}\text{Rb}/^{86}\text{Sr}$ against $^{87}\text{Sr}/^{86}\text{Sr}$ for some Allende refractory inclusions. ■, New values (see ref. 45): ALL 17 – $^{87}\text{Rb}/^{86}\text{Sr} = 0.00316$ – $^{87}\text{Sr}/^{86}\text{Sr} = 0.69900 \pm 0.0006$; ALL 15.4 – $^{87}\text{Rb}/^{86}\text{Sr} = 0.00714$ – $^{87}\text{Sr}/^{86}\text{Sr} = 0.69927 \pm 0.00014$; ALL JFM – $^{87}\text{Rb}/^{86}\text{Sr} = 0.0168$ – $^{87}\text{Sr}/^{86}\text{Sr} = 0.69990 \pm 0.00006$. ○, Data from the literature (refs 19–23). Since Allende refractory inclusions have been disturbed, extrapolation from the actual data points to an initial ratio can be meaningless. The Allende initial ratio (ALL) is obtained from a measured value. The $^{87}\text{Sr}/^{86}\text{Sr}$ initial values BABI^{17,18} and ADOR⁴¹ are indicated as well the error envelope of the isochron from Fig. 4.

single isochron of age $(4,520 \pm 40)$ Myr and strontium initial ratio 0.7003 ± 0.0004 .

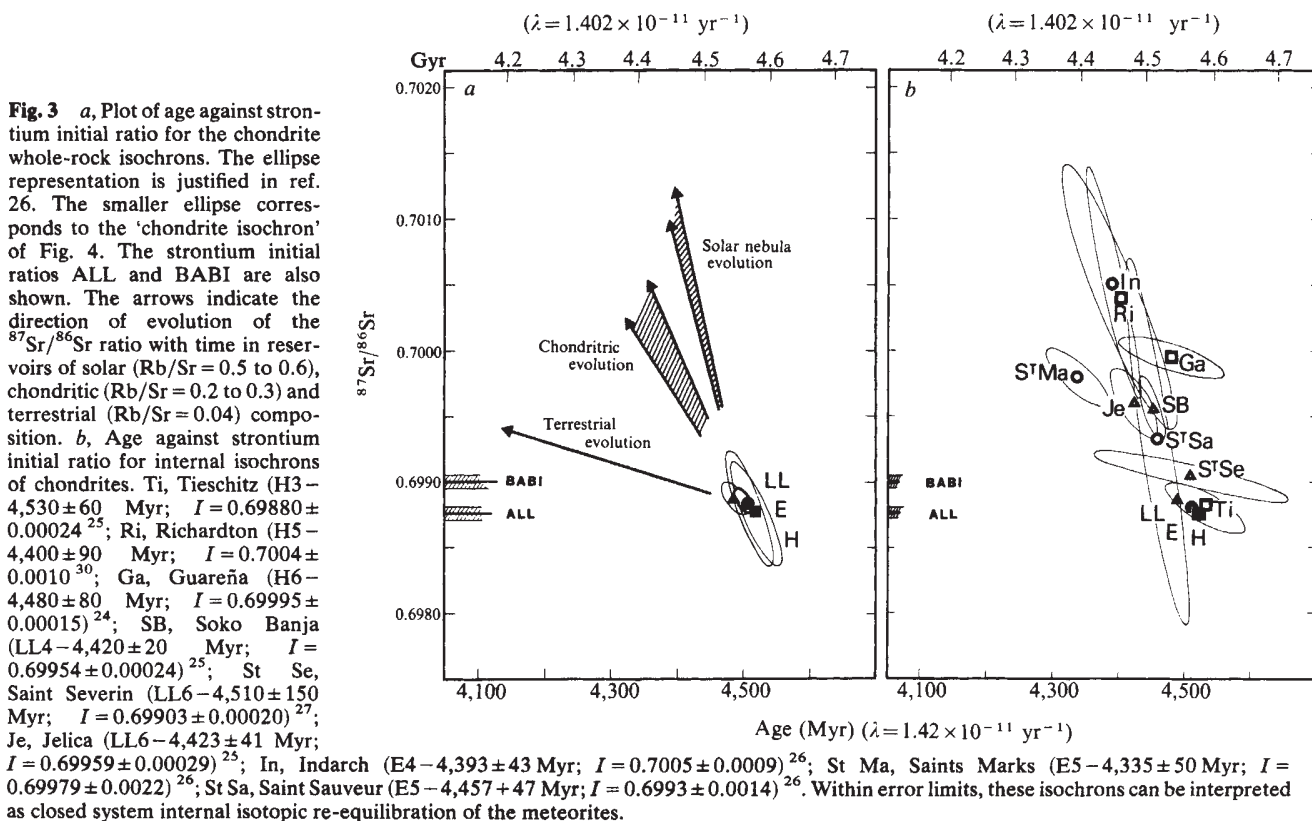
At the same time, Wasserburg and his co-workers demonstrated that it was possible to measure the isotopic composition of strontium at a precision of better than 10^{-4} (ref. 16) and obtained a very well-defined strontium initial ratio for basaltic achondrites of $^{87}\text{Sr}/^{86}\text{Sr}_{\text{BABI}} = 0.69899 \pm 0.00005$ (ref. 17, see also ref. 18). They have also proposed the use of the strontium initial ratio as a cosmochronometer. The principle of this technique is (Fig. 1): due to the decay of ^{87}Rb , the $^{87}\text{Sr}/^{86}\text{Sr}$ isotopic ratio increases with time in any reservoir. The larger the Rb/Sr

ratio of the reservoir, the faster this increase. The difference between the strontium initial ratio of two objects formed from a same reservoir gives the time elapsed between their formation, provided the Rb/Sr ratio of the reservoir is known. Basaltic achondrites are differentiated, that is, have suffered magmatic events after their formation; thus it could be expected that their initial ratio should be higher than that for chondrites. Indeed, that ratio for the refractory inclusions of the carbonaceous chondrite Allende was found to be $^{87}\text{Sr}/^{86}\text{Sr}_{\text{ALL}} = 0.69877 \pm 0.00005$ (refs 19–21; also refs 22, 23 and Fig. 2). Yet the chondrite isochron of Wetherill¹⁵, some precise data for low-radiogenic chondrites²⁰ as well as an internal isochron for the chondrite Guareña²⁴ indicated strontium initial ratios higher than BABI. Clearly more data with an improved accuracy were needed.

Since chondrites are radiogenic (the mean $^{87}\text{Sr}/^{86}\text{Sr}$ ratio for chondrites is about 0.75), the error on chondrite isochrons is dominated by that on the abscissa $^{87}\text{Rb}/^{86}\text{Sr}$. We developed a technique for measuring the rubidium isotopic composition with a reproducibility of better than 0.2% allowing a precision of $\pm 0.5\%$ for the $^{87}\text{Rb}/^{86}\text{Sr}$ ratio²⁵. This represented an increase of the precision for this ratio by a factor of 2–4 compared with the previous studies; joined with a better than 10^{-4} precision for the strontium isotopic composition¹⁸, this allowed a re-analysis of the problem. Specifically, we wanted to answer the following questions:

- (1) Do chondrites still define ^{87}Rb – ^{87}Sr isochrons with this increased precision?
- (2) Are there differences between the different groups of chondrites?
- (3) What is the strontium initial ratio for chondrites?
- (4) What is the ^{87}Rb – ^{87}Sr age for chondrites?

The results for the whole rock and mineral studies of the different groups of ordinary and enstatite chondrites have been published recently^{25–27}. However, the joint results of these group studies as well as their discussion have not yet been presented and are given here, with a new set of analyses for Allende white inclusions, carried out in our laboratory. Most of the recent results from other laboratories on ^{87}Rb – ^{87}Sr dating of chondrites will be incorporated into this discussion.



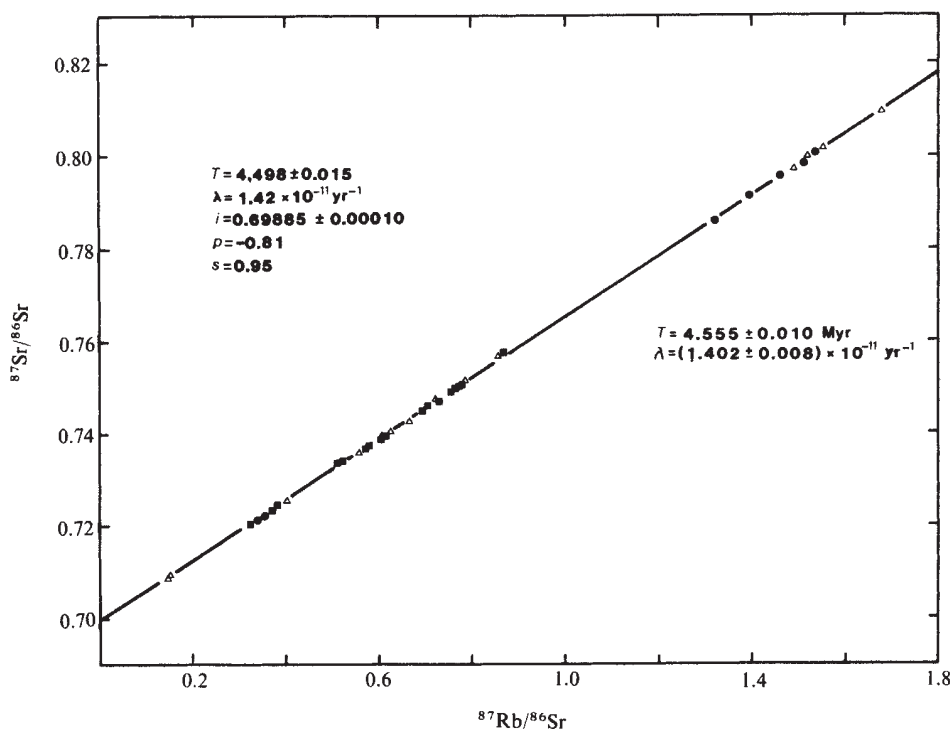


Fig. 4 Plots of $^{87}\text{Rb}/^{86}\text{Sr}$ against $^{87}\text{Sr}/^{86}\text{Sr}$ diagram for whole rock chondrites. Only these objects from the H(■), LL(△) and E(●) chondrites used in the previous isochron calculations are plotted and used in the least-squares regression. If the age is assumed to be of $4,555 \pm 10$ Myr, then $^{87}\text{Rb} = 1.402 \pm 0.008 \times 10^{-11} \text{ yr}^{-1}$.

Comparative chronology of chondrites

(1) Do chondrites still define whole-rock isochrons when the precision of the measurements is increased?

The answer to this question is positive: with a few exceptions, H, LL, and E-type chondrites define very precise isochrons. On the other hand, L chondrites scatter in the $^{87}\text{Rb}/^{86}\text{Sr}$ against $^{87}\text{Sr}/^{86}\text{Sr}$ diagram²⁵. Our preferred interpretation is that Rb has been lost by volatilization from these meteorites during the severe shocks which affected most of them: those which are unshocked tend to plot close to a reference isochron of age 4,500 Myr, whereas most of the shocked and degassed L chondrites plot on the left of this line, possibly indicating a lowered Rb/Sr ratio compared with their initial one.

A similar interpretation could explain most of the exceptions for the other groups of chondrites: for all of these objects, it is possible to find evidence of shock, brecciation or terrestrial contamination or alteration²⁵⁻²⁷. Such secondary disturbances could also explain the high initial ratios found for some low radiogenic chondrites by Wetherill *et al.*²⁰, as well as the high initial ratio for the chondrite isochron of Wetherill¹⁵. In what follows these 'disturbed' objects will be rejected. Note here that for H, LL and E chondrites, these amount to a small fraction of the analysed meteorites, and that they generally plot by more than 3σ off the line defining the others.

(2) Are there differences between the different groups?

The ages for the other groups were respectively:

H chondrites $T = 4,518 \pm 39$ Myr:

$$^{87}\text{Sr}/^{86}\text{Sr}_i = 0.69876 \pm 0.00040$$

LL chondrites $T = 4,486 \pm 20$ Myr:

$$^{87}\text{Sr}/^{86}\text{Sr}_i = 0.69887 \pm 0.00012$$

E chondrites $T = 4,508 \pm 37$ Myr:

$$^{87}\text{Sr}/^{86}\text{Sr}_i = 0.69880 \pm 0.00044$$

As noted by Sanz and Wasserburg²⁸, the errors on the age and initial ratio are anticorrelated. Our method²⁶ of calculating these correlation coefficients, gives -0.96 , -0.73 and -0.80 . We have also shown that, in a diagram where one plots the initial ratio against the age, an ellipse is always a reasonable representation of the error limits. The error ellipses for the

above isochrons are represented in Fig. 3. They are clearly undistinguishable from each other; in other words, the ^{87}Rb - ^{87}Sr age of all chondrites is identical within error limits.

(3) What are the internal isochrons of chondrites with respect to those for whole rocks?

It might be argued that to be sure that the different analysed objects be cogenetic, the best method is to analyse separate mineral phases of a given meteorite. Quite a few ^{87}Rb - ^{87}Sr internal studies of chondrites have now been achieved. These are Allende (CV)¹⁹⁻²³ (and this work), Tieschitz (H3)²⁵, Lost City (H5)²⁹, Richardton (H5)³⁰, Guareña (H6)²⁴, Mezö Madaras (L3)²⁵, Björbole (L4)³¹, Orvinio (L6)¹³, Farmington (L6)¹³, Bruderheim (L6)³², Chainpur (LL3)²⁵, Soko Banja (LL4)²⁵, Olivenza (LL5)²⁸, Gudder (LL5)²⁵, Krähenberg (LL5)³³, Peace River (LL6)^{32,19}, Saint Séverin (LL6)²⁷, Saint Mesmin (LL5)²⁵, Ensisheim (LL6)²⁵, Jelica (LL6)²⁵, Abee (E4)^{32,26}, Indarch (E4)^{13,26}, Saint Sauveur (E5)²⁶, Saint Mark's (E5)²⁶.

For some, the data obviously present analytical difficulties. Some do not have the necessary precision, number of analyses or range of the values in the ^{87}Rb - ^{86}Sr , ^{87}Sr - ^{86}Sr diagram, to be useful for comparative chronology. In addition, most of the remaining results do not define an isochron. In most cases, this can result of shock resetting or terrestrial alteration. We have presented in Fig. 3 the few which we consider reliable and useful. The obvious result is that only the H3 chondrite Tieschitz has the same internal age as that for whole rocks. All others have a younger age and a higher $^{87}\text{Sr}/^{86}\text{Sr}$ initial ratio. Note that within error limits, all these internal isochrons are consistent with a closed system reequilibration of the meteorites. Because Tieschitz is an unequilibrated object, whereas most others are equilibrated, this is interpreted as being due to the cooling or metamorphism history of the latter²⁵⁻²⁷. As for E-chondrite, this could reveal shock effects²⁶. From the $^{87}\text{Sr}/^{86}\text{Sr}$ initial ratio of these internal isochrons, it is possible to estimate the interval of time elapsed between the chondrite formation and internal isotopic homogenization¹⁷. For the type 6 objects of Fig. 3, this interval is of the order of 100 Myr. It is shorter, of the order of 40 Myr, for the type 4 object Soko Banja.

The internal isochron for Tieschitz is a 'chondrule isochron'. This means that it has not been obtained by analysing pure mineral phases, but whole chondrules²⁵. This result shows that, at the precision limits of the method, chondrules were formed simultaneously with whole rock chondrites.

The whole rock best isochron for chondrites

Comparison of the ages of the various chondrite classes shows that they are undistinguishable within error limits. It is thus justified to use all data together and determine a ^{87}Rb - ^{87}Sr 'best' age for chondrites. The results are shown in Figs 4 and 5: the best age for chondrites is $4,498 \pm 15$ Myr for $\lambda^{87}\text{Rb} = 1.42 \times 10^{-11} \text{ yr}^{-1}$.

Let us consider the consequences of this result.

The precision for this age is the 'best' ever obtained for a ^{87}Rb - ^{87}Sr isochron; it is comparable to that for extinct radioactivities such as ^{129}I - ^{129}Xe (see ref. 34), but relates to an absolute age. Since the scatter of the data points around the isochron is small (the reduced X^2 is 0.95), this error limit is probably significant. On Fig. 5, we have represented the ^{87}Rb - ^{87}Sr model age of the various chondrites relative to the $^{87}\text{Sr}/^{86}\text{Sr}$ initial ratio of the isochron from Fig. 4. Most results indeed agree with the best fit line within error limits: 95% of the data points plot within an interval of ± 50 Myr. More than 80% plot within an interval of ± 30 Myr. This result confirms the idea of a 'sharp isochronism' of formation of chondrites, on a time scale of a few million years as deduced from I-Xe chronology³⁴.

The question is then to relate this age to the processes of formation of the Solar System. The isochron is based on variations of the Rb/Sr ratio among chondrites. As the Sr concentration for the objects of a given group is constant³⁵ (Fig. 6), the variations are due to variations of the Rb concentration. It is generally believed that the latter is either the result of fractionation of moderately volatile elements during condensation-accretion, or of volatilization and loss during the reheating and metamorphic events which affected most chondrites (see ref. 36). ^{87}Rb - ^{87}Sr dating does not help to decipher which process is being dated. However, in order not to perturb the whole rock isochron, the latter process should have occurred during the first few million years of the chondrites. In particular it is unlikely that Rb was lost some 100 Myr after the formation of these meteorites, at the time dated by internal ^{87}Rb - ^{87}Sr isochrons of equilibrated chondrites. Rb is also fractionated among various clasts of a given chondrite and it is not generally known when such clasts assembled. If it could be demonstrated that their assembly occurred before metamorphism, this would be a strong evidence in favour of Rb fractionation during condensation-accretion.

It is known that Rb is highly fractionated in the Earth compared with ordinary chondrites: from the Sr-Nd isotopic correlation in terrestrial basalts, it is estimated that the 'bulk Earth' $^{87}\text{Rb}/^{86}\text{Sr}$ ratio is 0.098 (refs 37, 38). Not that only Ensisheim presents such a low value. It is not certain whether the fractionation processes which affected moderately volatile elements were the same inside the chondrite parent bodies as they were in a whole terrestrial planet. Yet the present results indicate that fractionation could attain the same efficiency on very small to very large distance scales. Identification of these processes could result in very strong constraints for the chronology of formation of the Earth.

The best age for chondrites is different from the most recent ^{206}Pb - ^{207}Pb ages for meteorites. The latter can be very precise because they rely on isotopic ratios only. However, their actual values are very sensitive to small disturbances³⁹. If one selects the data for which the U-Pb or U-Th-Pb ages are concordant, which means that the ages deduced from the independent chronometers ^{235}U - ^{207}Pb , ^{238}U - ^{206}Pb and ^{232}Th - ^{208}Pb are equal, the corresponding ^{206}Pb - ^{207}Pb ages fall around $4,555 \pm 10$ Myr, whether for whole rocks chondrites³⁹, for chondritic minerals²⁷, or for achondrites⁴¹. We do not intend to discuss the problems of U-Pb dating of chondrites fully here but it is possible that the actual U-Pb age of some chondrites is as old as 4,580 Myr (ref. 39). In any case, even the 4,555 Myr value is 1% older than our ^{87}Rb - ^{87}Sr age. It seems unrealistic that such a difference is real, and that Pb-Sr and U-Th-Pb fractionations occurred independently on a time scale of about 50 Myr.

As all our measurements were made using the same method, and the same spike solution²⁵⁻²⁷, such a difference could be due to a systematic bias in our technique. The main possible source for such a bias would be the spike calibration. However, our spike has been repetitively calibrated against different normal solutions and the results agreed within 0.2%. In addition, these spike solutions were used to analyse the Rb and Sr concentrations in NBS glass wafers and our results agreed within 0.2% of the certified values. Finally, our values agree very well with those obtained by others³⁰.

On the whole, we believe that a 1% bias in our measurements is unreasonably large. A possible alternative is that the ^{87}Rb decay constant is still incorrectly estimated. Our results compared with the ^{206}Pb - ^{207}Pb ages would indicate that $\lambda^{87}\text{Rb} = (1.402 \pm 0.008) \times 10^{-11} \text{ yr}^{-1}$. Compared with the value of $1.42 \times 10^{-11} \text{ yr}^{-1}$, this would change ^{87}Rb - ^{87}Sr ages by a factor of 1.013. This should be verified by comparison of other simultaneous Rb-Sr, Sm-Nd and Pb-Pb measurements; however, such studies should attain a better than 1% precision for eventual differences to be significant.

Note that K-Ar ages of chondrites are also younger than U-Th-Pb ages: Turner *et al.*⁴⁰ found an average ^{39}Ar - ^{40}Ar plateau age of $4,480 \pm 30$ Myr for such objects, the older one being of 4,520 Myr. Their interpretation is that the K-Ar system of the bulk meteorites has been reset during metamorphism on a time scale of 100 Myr. If such was the case, one would expect that Rb should also be redistributed simultaneously during this process but, as discussed above, this seems difficult. In addition, no systematics of ^{39}Ar - ^{40}Ar ages against the degree of metamorphism is found. Alternatively, a bias in the ^{40}K decay constant joined by larger than estimated error limits would render the problem simpler to understand. Again, these questions warrant more data (see the discussion in ref. 25).

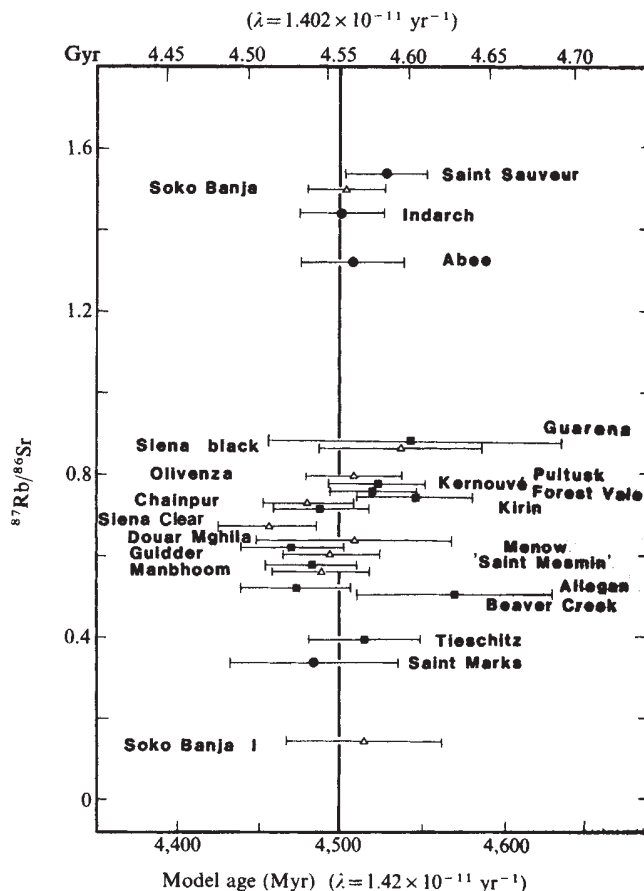


Fig. 5 Model ^{87}Rb - ^{87}Sr ages of chondrites. The $^{87}\text{Sr}/^{86}\text{Sr}$ initial ratio is assumed to be that of our best fit isochron (Fig. 4). The different analyses for a given object have been averaged. All ordinary and enstatite chondrites have undistinguishable ages.

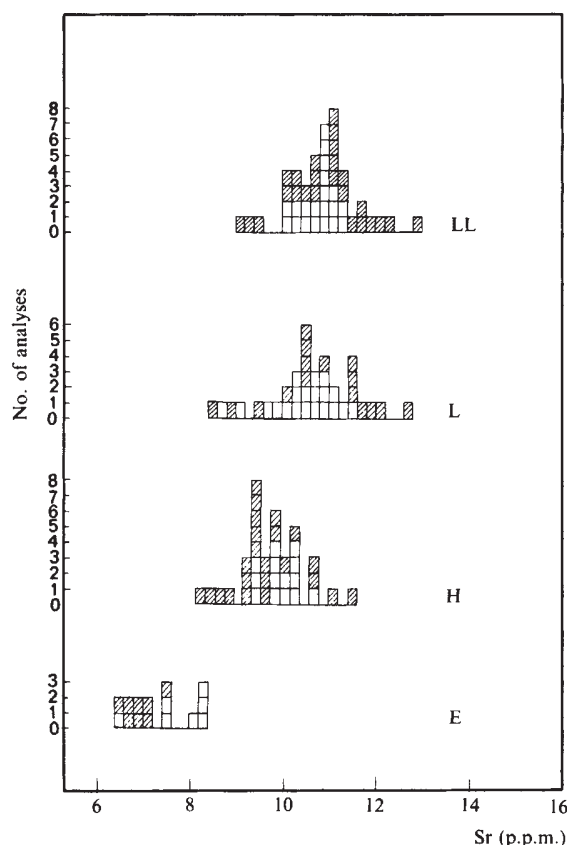


Fig. 6 Sr concentration in ordinary and enstatite chondrites. Our data^{25,26} (▨) and those of Gopalan and Wetherill³⁵ (□) are included. Only objects which plot close to the Rb–Sr isochron have been considered. Much of the scatter is probably due to the non-representativity of the samples. The Sr concentration within each group is constant.

The $^{87}\text{Sr}/^{86}\text{Sr}$ initial ratio for chondrites is 0.69885 ± 0.00010 . Figures 2 and 3 also show $^{87}\text{Sr}/^{86}\text{Sr}_{\text{BABI}}$ and $^{87}\text{Sr}/^{86}\text{Sr}_{\text{ALL}}$, where $^{87}\text{Sr}/^{86}\text{Sr}_{\text{BABI}}$ is the initial ratio for basaltic achondrites¹⁷, and $^{87}\text{Sr}/^{86}\text{Sr}_{\text{ALL}}$ is that for Allende white inclusions¹⁹. The value for basaltic achondrites has been confirmed in our laboratory¹⁸. We also show in Fig. 2 a few data on Allende white inclusions. We assume that no interlaboratory bias has been introduced in the strontium initial ratios for meteorites, and that for the 'chondrite isochron' is significantly lower than BABI and undistinguishable from ALL. Thus, the expected result that the differentiated achondrites have a strontium initial ratio higher than that of chondrites is confirmed.

Yet, from Figs 2 and 3, it might appear that ALL is slightly lower than the chondrite initial ratio. The actual value for chondrites is very close to that observed in the differentiated meteorite Angra Dos Reis (Ador, see for example ref. 41). Such a difference could be due to a time delay in the chondritic fractionation of Rb relative to the formation of the Allende refractory inclusions. Assuming that both formed directly from the solar nebula, the Rb/Sr ratio of which is 0.6 to 0.5 (ref. 42), the time delay would be 3.0 ± 5.6 Myr to 3.9 ± 7.3 Myr. If the Rb fractionation which is being dated by the chondrite isochrons occurred during chondrite metamorphism, these objects would have evolved in a reservoir of chondritic composition ($\text{Rb}/\text{Sr} = 0.3$ to 0.2) and the time delay would be 7.0 ± 13.2 to 9.4 ± 17.6 Myr. Thus, chondrites and Allende refractory inclusions are 'synchronous' on a time scale of 10 Myr. This time scale is comparable with that deduced from I–Xe chronology³⁴. Unfortunately, it is not precise enough for constraining the models of formation of planetesimals (see ref. 43), since all these models deal with time scales shorter than 10 Myr.

Another indirect result of the existence of an isochron is that there was no large anomaly for the ^{87}Sr and ^{86}Sr isotopic

abundances at the time of formation of chondrites in the areas of the solar nebula where these meteorites formed. Evidence has been found for isotopic abundance anomalies of ^{84}Sr in some rare refractory inclusions of the Allende carbonaceous chondrite⁴⁴. It is not clear whether there were some anomalies for ^{87}Sr and ^{86}Sr in this material. For whole rock chondrites, the present evidence is that isotopic anomalies for these isotopes were small enough for the ^{87}Rb – ^{87}Sr dating technique to be fully useable.

Conclusions

The major conclusion of our study is that all chondrites formed simultaneously from a solar nebula where the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio was homogeneous; the $^{87}\text{Sr}/^{86}\text{Sr}$ initial ratio for chondrites is undistinguishable from that for carbonaceous chondrites.

This ratio is significantly lower than BABI, the initial ratio for basaltic achondrites^{17,18}. This can be interpreted as a later differentiation of the parent body of these latter objects. However, the matter is complicated by the fact that there exists at least one differentiated object with an initial ratio lower than BABI and undistinguishable from that for chondrites (ADOR, ref. 41). In addition, ^{87}Rb – ^{87}Sr dating does not give any information about the predifferentiation history of these bodies (see ref. 18). The same applies for the Earth and the Moon. Yet, our results indicate that use of the ALL value of 0.69877 ± 0.00005 (ref. 19) as a starting value for the history of all these planets is not unpalatable.

Using the ^{87}Rb decay constant of $\lambda = 1.42 \times 10^{-11} \text{ yr}^{-1}$, the ^{87}Rb – ^{87}Sr age for chondrites is $4,498 \pm 15$ Myr. For the Rb–Sr and U–Th–Pb chronometers to agree with each other at 4,555 Myr, this decay constant would be $1.402 \pm 0.008 \times 10^{-11} \text{ yr}^{-1}$. This result calls for further simultaneous precise Rb–Sr, Sm–Nd and U–Th–Pb analyses of selected samples. It should certainly be considered in the comparison of precise ages obtained by different chronometers.

^{87}Rb – ^{87}Sr dating of chondrites relates to Rb fractionation in the early Solar System. This occurred either during condensation–accretion or at an early stage of metamorphism of chondrites. The ^{87}Rb – ^{87}Sr internal system has been perturbed in most chondrites, probably due to shock and thermal metamorphism: type 3 unequilibrated chondrites are an exception and can be considered unique.

We thank G. Manhes and all our colleagues for discussions, the technicians of the laboratory for maintenance of machines, P. Pellas and numerous curators for providing the samples, and D. Unruh and G. Wetherill for helpful criticisms.

Received 13 May; accepted 11 October 1982.

- Steiger, R. H. & Jäger, E. *Earth planet. Sci. Lett.* **36**, 359–362 (1977).
- Neumann, W. & Huster, E. *Z. Phys.* **270**, 121–127 (1974); *Earth planet. Sci. Lett.* **33**, 277–288 (1976).
- Aldrich, L. T., Wetherill, G. W. & Tilton, G. R. *Phys. Rev.* **103**, 1045–1047 (1956).
- Jaffey, A. H., Flynn, K. F., Glendenin, L. E., Bentley, W. C. & Essling, A. M. *Phys. Rev. C*, **4**, 1899–1906 (1971).
- Schumacher, E. *Z. Naturforsch.* **II**, **20**, 206–212 (1956).
- Herzog, L. F. & Pinson, W. H. *Am. J. Sci.* **254**, 555–566 (1956).
- Webster, R. K., Morgan, J. W. & Smales, A. A. *Trans. Am. geophys. Un.* **38**, 543–545 (1957).
- Gast, P. W. *Geochim. cosmochim. Acta* **26**, 927–943 (1962).
- Pinson, W. H., Schnetzler, C. C., Beiser, E. E., Fairbairns, H. M. & Hurley, P. M. *Geochim. cosmochim. Acta* **29**, 455–466 (1965).
- Urey, H. C. & Craig, H. *Geochim. cosmochim. Acta* **4**, 36–82 (1953).
- Clayton, R. N., Onuma, N. & Mayeda, T. K. *Earth planet. Sci. Lett.* **30**, 10–18 (1976).
- Murthy, V. R. & Compston, W. *J. geophys. Res.* **70**, 5297–5307 (1965).
- Gopalan, K. & Wetherill, G. W. *J. geophys. Res.* **73**, 7133–7136 (1968); **74**, 4349–4358 (1969); **75**, 3457–3467 (1970); **76**, 8484–8492 (1971).
- Kaushal, S. K. & Wetherill, G. W. *J. geophys. Res.* **74**, 2717–2726 (1969); **75**, 463–468 (1970).
- Wetherill, G. W. *Ann. Rev. nucl. Sci.* **25**, 283–328 (1975).
- Wasserburg, G. J., Papanastassiou, D. A., Menow, E. V. & Bauman, C. A. *Rev. Scient. Instrum.* **40**, 288–293 (1969).
- Papanastassiou, D. A. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **5**, 361–376 (1969).
- Birck, J.-L. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 37–51 (1978); **26**, 29–35 (1973); **55**, 116–122 (1981); *Nature* **282**, 288–289 (1979).
- Gray, C. M., Papanastassiou, D. A. & Wasserburg, G. J. *Icarus* **20**, 213–239 (1973).
- Wetherill, G. W., Mark, R. & Lee Hu, S. *Science* **182**, 2781–2783 (1973).
- Nyquist, L. E., Hubbard, N. J. & Gast, P. W. in *Lunar Science Vol. 4* (eds Chamberlain, J. & Watkins, C.) 567–570 (Lunar Science Institute, Houston, 1973).
- Tatsumoto, M., Unruh, D. M. & Desborough, G. A. *Geochim. cosmochim. Acta* **40**, 617–634 (1976).
- Nagasawa, M. & Borming, J. in *Lunar Science Vol. 7*, 591–593 (Lunar Science Institute, Houston, 1976).

24. Wasserburg, G. J., Papanastassiou, D. A. & Sanz, H. G. *Earth planet. Sci. Lett.* **7**, 33–43 (1969).
25. Minster, J.-F. & Allègre, C. J. *Earth planet. Sci. Lett.* **32**, 191–198 (1976); **42**, 333–347 (1979); **56**, 89–106 (1981); *Meteoritics* **14**, 235–248 (1979); *Nature* **278**, 732–734 (1979).
26. Minster, J.-F., Ricard, L.-P. & Allègre, C. J. *Earth planet. Sci. Lett.* **44**, 420–440 (1979).
27. Manhès, G., Minster, J.-F. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 14–24 (1978).
28. Sanz, H. G. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **6**, 335–345 (1969).
29. Schnetzler, C. C. & Bottino, M. L. *J. geophys. Res.* **76**, 4061–4069 (1971).
30. Evensen, N. M., Carter, S. R., Hamilton, P. J., O'Nions, R. K. & Ridley, W. J. *Earth planet. Sci. Lett.* **42**, 223–236 (1979).
31. Shields, R. M., Pinson, W. H. & Hurley, P. M. *J. geophys. Res.* **71**, 2163–2167 (1966).
32. Shima, M. & Honda, M. *Earth planet. Sci. Lett.* **2**, 337–343 (1967).
33. Kempe, W. & Müller, O. in *Meteorite Research* (ed. Millman, P.) 418–428 (Reidel, Dordrecht, 1969).
34. Podosek, F. A. *Geochim. cosmochim. Acta* **34**, 341–365 (1970).
35. Gopalan, K. & Wetherill, G. W. *Handbook of Elemental Abundances in Meteorites*, 297–302 (ed. Mason, B.) (Gordon and Breach, New York, 1971).
36. Wasson, J. T. *Meteorites* (Springer, Berlin, 1974).
37. De Paolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 743–746 (1976).
38. Allègre, C. J., Ben Othman, D., Polve, M. & Richard, P. *Phys. Earth planet. Inter.* **19**, 293–306 (1979).
39. Manhès, G. & Allègre, C. J. *Meteoritics* **13**, 543–548 (1978); (in the press).
40. Turner, G., Enright, M. C. & Cadogan, P. H. *Proc. 9th Lunar Sci. Conf.* 989–1025 (1978).
41. Wasserburg, G. J., Tera, F., Papanastassiou, D. A. & Huneke, J. C. *Earth planet. Sci. Lett.* **35**, 294–326 (1977).
42. Ross, J. E. & Aller, L. H. *Science* **191**, 1223–1229 (1976).
43. Wetherill, G. W. *Ann. Rev. Astr. Astrophys.* **18**, 77–113 (1979).
44. Papanastassiou, D. A. & Wasserburg, G. J. *Geophys. Res. Lett.* **5**, 595–598 (1978).
45. Birck, J.-L., Lorin, J.-C. & Allègre, C. J. *Meteoritics* **12**, 179 (1977).

Human chorionic gonadotropin β -subunit is encoded by at least eight genes arranged in tandem and inverted pairs

William R. Boorstein*, Nikos C. Vamvakopoulos* & John C. Fiddes

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

The β -subunit of the human placental glycoprotein hormone, chorionic gonadotropin (HCG), is coded for by at least eight different genes or pseudogenes, four of which are arranged in inverted pairs while the other four are arranged in tandem pairs. The genes are each 1.45 kilobases (kb) long and have two short introns located in identical positions.

THE placental hormone, human chorionic gonadotropin (HCG), together with the three pituitary hormones, luteinizing hormone (LH), follicle-stimulating hormone (FSH) and thyroid-stimulating hormone (TSH), comprise the glycoprotein hormone family. All four hormones are dimeric with two different subunits, α and β , which are associated noncovalently (for review see ref. 1).

Amino acid sequence analysis has shown that the α -subunit seems to be common to the four hormones, while the β -subunits are unique and confer biological specificity on the individual hormones. Varying degrees of amino acid sequence homology exist between the four β -subunits, which led to the proposal that they evolved from a common ancestral gene¹.

We have previously isolated and characterized cDNA clones encoding the α - (ref. 2) and β - (ref. 3) subunits of the placental hormone, HCG. The α -subunit cDNA was then used as a hybridization probe to purify the corresponding genomic fragment⁴. Comparison of the structure of the cloned α -subunit gene with the hybridizing restriction enzyme fragments in human DNA led to the conclusion that there is only a single α -subunit gene which is transcribed to make the α -subunit mRNA for all four hormones⁴. A similar conclusion was also reached by Boothby *et al.*⁵, who used an α -subunit cDNA to probe uncloned human DNA.

Here we show that the gene structure for the β -subunit of HCG is much more complex than that of the single α -subunit gene. At least eight distinct β -HCG or β -HCG-like genes have been identified by a combination of filter hybridization of genomic DNA and isolation of bacteriophage λ genomic recombinants. These genes, which are structurally very similar, form a family of sequences which are probably all linked.

Isolation of β -HCG genomic recombinants

The β -HCG cDNA was used to isolate genomic recombinants from two independent human chromosomal libraries, one in

bacteriophage λ Charon 4A (ref. 6) and one in bacteriophage λ Charon 28 (E. Fritsch and C. Sabourin, unpublished results). We purified 16 hybridizing phage, 9 from the Charon 4A library and 7 from the Charon 28 library. Restriction enzyme maps for these phage DNAs were determined by analysing a combination of single and double restriction enzyme digests using the filter hybridization method of Southern⁷. We thus identified three different groups of overlapping recombinants (the restriction enzyme maps for these three groups, *a*, *b* and *c*, are shown in Fig. 1).

β -HCG genes are arranged in both inverted and tandem pairs

Eight separate β -HCG, or β -HCG-like, genes or pseudogenes have been identified by hybridization in the three groups, *a*, *b*, and *c*; the location and polarity of these genes, termed β -HCG 1–8, are shown in Fig. 1.

Four of the genes, β -HCG 1 and 2 and β -HCG 5 and 6, are found in inverted pairs. A detailed map of the region containing the gene pair β -HCG 5 and 6 is shown in Fig. 2; DNA sequencing of the subcloned fragments identified in this figure confirmed the inverted orientation of the genes. Similarities between the restriction enzyme patterns of these β -HCG 5 and 6 subclones with similar subclones containing β -HCG 1 and 2 showed that these two genes also form an inverted pair.

In contrast to these four genes, β -HCG 3 and 4 and β -HCG 7 and 8 form tandem gene pairs. We have shown this again by comparing restriction enzyme digestion patterns of subcloned fragments and, in the case of β -HCG 3, by DNA sequencing.

β -HCG genes contain two introns in identical positions

DNA sequencing of four of the eight β -HCG genes (β -HCG 2, 3, 5 and 6; K. Talmadge, N.C.V., W.R.B. and J.C.F., unpublished results) reveals that they all have two introns that separate codon –16 of the signal peptide from codon –15, and codon 41 of the mature protein from codon 42. The two introns,

* Present addresses: Department of Molecular Biology, University of Wisconsin, Madison, Wisconsin 53706, USA (W.R.B.); 28 Kononos, Pagrati, Athens, Greece (N.C.V.).

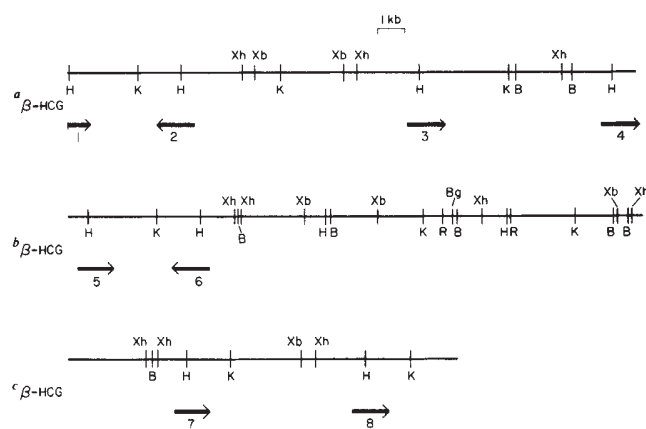


Fig. 1 Restriction enzyme maps of the three groups of β -HCG genomic isolates (a, b and c). H = *Hind*III, K = *Kpn*I, Xh = *Xho*I, Xb = *Xba*I, B = *Bam*HI, Bg = *Bgl*II and R = *Eco*RI. The map for group a (21.5 kb long) was constructed from 10 overlapping phage isolates, that for group b (22.0 kb) from 5 isolates and that for c (14.7 kb) from a single isolate. The positions of the eight β -HCG genes (1–8) are indicated by arrows which point in the proposed direction of transcription. Each gene spans ~ 1.45 kb from the cap site to the poly(A) site. These phage isolates were obtained either from a genomic library made from human fetal liver DNA, partially digested with *Hae*III and *Alu*I, and cloned with *Eco*RI linkers into bacteriophage λ Charon 4A (a gift from T. Maniatis⁶) or from a second library, made from DNA isolated from the chemically transformed human fibroblast cell line, HAI, and cloned as an *Mbo*I partial digest in the vector bacteriophage λ Charon 28 (a gift from E. Fritsch). These libraries, and the individual bacteriophage isolates, were propagated in *E. coli* K803 in P1 conditions in accordance with the NIH Guidelines for Recombinant DNA Research. The 579-base pair (bp) *Hind*III fragment containing the cloned β -HCG cDNA isolated previously³ was ³²P-labelled either by nick-translation¹² or by priming with calf thymus DNA oligonucleotides¹³. Recombinant phage plaques were screened on Schleicher and Schuell 0.45- μ m nitrocellulose filters by the method of Benton and Davis¹⁴ as described elsewhere⁴. Restriction endonucleases were from New England Biolabs. Digests of DNA isolated from recombinant bacteriophage were analysed by electrophoresis on 1.0% agarose gels, followed by transfer to nitrocellulose filters and hybridization by the method of Southern⁷ as described elsewhere⁴. The sizes of fragments less than 2 kb long were estimated on 3% acrylamide gels.

named A and B, are ~ 350 and 230 nucleotides long, respectively. Figure 3 shows a composite structure for these genes. As the other genes, β -HCG 1, 4, 7 and 8, have restriction enzyme patterns very similar to the partially sequenced β -HCG genes, we expect them also to have two introns at the same positions.

The *Hind*III site in intron A (Fig. 3) is the *Hind*III site at the end of the 2.6- and 1.6-kb *Hind*III-*Kpn*I fragments identified in Fig. 2. These fragments, which contain β -HCG 5 and 6 respectively, were subcloned into plasmid vectors. The 5' exon and the promoter region for β -HCG 5 and 6 are contained on a 0.8-kb *Hind*III to *Eco*RI fragment and a 1.4-kb *Hind*III to *Bam*HI fragment, respectively (Fig. 2). These fragments were also isolated from a phage λ recombinant and subcloned into plasmids. The DNA sequencing used to generate the composite map in Fig. 3 was obtained from these subclones.

The 0.8-kb *Hind*III to *Eco*RI fragment of Fig. 2 was also used as a hybridization probe in filter hybridization analysis of all the phage λ clones. This fragment, which contains the 5' end of β -HCG 5, hybridizes strongly to the region upstream from the *Hind*III site observed in all the β -HCG genes. This information, together with the detailed restriction enzyme analysis of the region upstream from these *Hind*III sites, further supports our conclusion that all eight β -HCG genes have the

same general structure with two intervening sequences in identical positions. Small differences between the genes, and the possibility that there may be pseudogenes, cannot be excluded until the DNA sequence analysis is completed.

Correlation of cloned fragments with those in total genomic DNA

To determine whether the cloned β -HCG genes correlate with the hybridizing fragments seen in total genomic DNA and to estimate the number of β -HCG hybridizing genes or pseudogenes, we digested human placental DNA with several restriction endonucleases, followed by filter hybridization analysis⁷. Two separate probes were used: the β -HCG cDNA specific for the region to the 3' side of the *Hind*III site located in intron A (Fig. 3), and the 0.8-kb *Hind*III to *Eco*RI fragment specific for the region to the 5' side of this site (Fig. 2). An autoradiograph of a filter hybridization of genomic DNA with the 3' probe is shown in Fig. 4 while one with the 5' probe is shown in Fig. 5. Both of these probes identify a fairly complex pattern of hybridizing fragments. This is consistent with the conclusion, drawn from the analysis of the phage λ recombinants, that there are multiple regions hybridizing to the β -HCG cDNA. The intensity differences among the hybridizing bands make it difficult to estimate accurately the number of β -HCG sequences.

Examination of the restriction enzyme maps of Fig. 1 suggests that in a *Kpn*I digest of genomic DNA, each gene should be contained in a unique fragment that is recognized by both the 5' and the 3' probes. Similarly, in a combined *Hind*III-*Kpn*I digest, the genes should be separated into two parts that are recognized exclusively by either probe. The hybridization patterns observed in the filter hybridizations of human genomic DNA are consistent with these expectations. With *Kpn*I alone, both probes (Figs 4, 5) identify the same set of bands, but, as a result of intensity differences, it is hard to estimate the number of genes represented by these bands. With the combined *Hind*III-*Kpn*I digests, both probes (Figs 4, 5) give different, but rather simpler, patterns. For example, in the case of the 3' probe with a *Hind*III-*Kpn*I digest (Fig. 4), four bands of 3.5, 2.6, 2.2 and 1.6 kb are identified. Examination of restriction enzyme maps of the cloned genes (Fig. 1) reveals that β -HCG genes 1 and 5 are both on 2.6-kb *Hind*III-*Kpn*I fragments, β -HCG 2, 6, 7 and 8 are on 1.6-kb fragments, β -HCG 3 is on

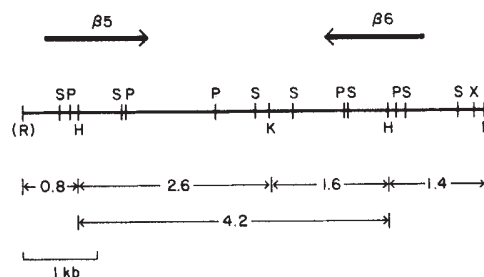


Fig. 2 Detailed restriction enzyme map for the region of group b containing the symmetrical pair of genes $\beta 5$ and $\beta 6$. The polarity of these genes is indicated by arrows. Restriction enzyme sites are shown for *Eco*RI (R), *Hind*III (H), *Kpn*I (K), *Bam*HI (B), *Sac*I (S), *Xho*I (X) and *Pst*I (P). The *Eco*RI site shown in parentheses is the result of linker ligation in the construction of the Charon 4A library. This map was obtained from an analysis of the following subcloned fragments whose positions are indicated: *Eco*RI-*Hind*III, 0.8 kb; *Hind*III-*Kpn*I, 1.6 and 2.6 kb; *Bam*HI-*Hind*III, 1.4 kb; and *Hind*III-*Hind*III, 4.2 kb. Double cut *Hind*III/*Eco*RI and *Hind*III/*Bam*HI vectors were prepared from pBR322 while a *Hind*III/*Kpn*I vector was prepared from pSM804, a pBR322 derivative containing a unique *Kpn*I site (obtained from E. Stavnezer). Subcloning of the λ DNA fragments was by standard methods using *E. coli* DH1 (D. Hanahan, unpublished results) as the host strain.

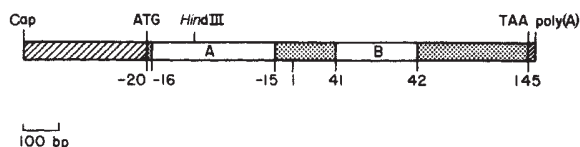


Fig. 3 Composite structure of β -HCG gene. Dotted areas represent coding regions while the cross-hatched areas represent the 5' and 3' untranslated regions. The positions of the two introns (A and B) are shown. The locations of the predicted cap site (Cap), the initiation codon (ATG), termination codon (TAA) and the polyadenylation site [poly (A)] are marked. Numbers at the bottom of the diagram refer to amino acids (negative numbers refer to the signal peptide). The location of the single *HindIII* site referred to in the text is shown. DNA fragments were labelled at their 3' termini by DNA polymerase I (Klenow fragment; New England Biolabs) with [α - 32 P]-labelled dNTPs (Amersham). The sequencing method of Maxam and Gilbert⁵ was used.

a 3.5-kb fragment, while β -HCG 4 is on a fragment of unknown size as it is at the end of a phage λ clone. There is no evidence in the cloned isolates of the hybridizing 2.2-kb fragment seen in the filter hybridizations of genomic DNA.

Similarly, with the 5' probe (Fig. 5) there are *HindIII*-*KpnI* bands at 5.0, 4.0 and 3.7 kb. Again this correlates well with the map of the cloned genes. The 5' ends of β -HCG genes, 2 and 4 are on 3.7-kb fragments, those of β -HCG 3, 6 and 8 are on 5.0-kb fragments, while those of β -HCG 1, 5 and 7 are on fragments of unknown size. As with the 2.2-kb fragment seen with the 3' probe, there is no evidence in the cloned isolates of a 4.0-kb fragment hybridizing to the 5' probe. Both of these missing fragments are seen in the filter hybridization of genomic DNA as weakly hybridizing bands and are thus good candidates for containing β -LH sequences which have not yet been detected in the phage library screening (see below).

The apparent simplicity of the filter hybridization of genomic DNA with the combined *HindIII*-*KpnI* digests when compared with *KpnI* alone is a reflection of the duplication of sequences observed in the β -HCG locus. For example, the 1.6-kb *HindIII*-*KpnI* band which hybridizes strongly to the 3' probe (Fig. 4) is known from the analysis of the bacteriophage recombinants, to contain at least four genes— β -HCG, 2, 6, 7 and 8.

The symmetrical organization of the β -HCG genes presents additional problems in correlating the cloned fragments with the bands observed in the genomic filter hybridizations. For example, after hybridization of the 3' probe to a *HindIII* digest (Fig. 4), the expected 4.2-kb band is observed but at a much lower intensity than expected, as we have identified two distinct 4.2-kb fragments each of which contain two genes. In other filter hybridizations of genomic DNA (data not shown), we observed variability in the intensity of this band and attribute this to the ability of the 4.2-kb inverted fragments to form snap-back structures on renaturation during the transfer procedure. This would reduce the availability of sequences for hybridizing to the β -HCG cDNA probe and would therefore reduce the observed level of hybridization.

Because the DNA being used for the filter hybridization analysis and for the construction of the two genomic libraries came from different individuals, it is possible that the correlation of the filter hybridization data with the cloned fragments is confused by polymorphism. We have, however, found no evidence for this as the *KpnI* and *HindIII*-*KpnI* fragments that hybridize β -HCG from 10 different sources, including the DNA used to make the Charon 28 library (data not shown), are identical.

While all these factors make it difficult to determine the precise number of β -HCG-related sequences in the human genome, we believe that a reasonable estimate, based mainly on the bacteriophage λ clones we have isolated and on the *KpnI* and *HindIII* digests of human DNA, is nine, of which eight have been isolated and characterized.

Are all the β -HCG genes physically linked?

Although it is not known to which chromosome the β -subunit of HCG maps, we have some evidence for the linkage of all the β -HCG hybridizing fragments. A single hybridizing band is observed in filter hybridizations of human DNA digested with either *EcoRI* or *SalI* (our unpublished results). In the case of *EcoRI*, this fragment is between 50 and 60 kb long. Although we are unable to exclude the possibility that a band of this size contains several hybridizing fragments, which were not resolved by electrophoresis, it suggests that the β -HCG genes may all be linked. Consistent with a single hybridizing fragment is the fact that in the phage λ recombinants there are no *SalI* sites and the only *EcoRI* sites are those two observed upstream from the 5' end of β -HCG 6 (Fig. 1). The first of these sites may represent one end of a large *EcoRI* fragment which defines the β -HCG locus.

Despite evidence suggesting that the genes are all linked, extensive screening of two distinct genomic libraries with a β -HCG cDNA hybridization probe resulted in the isolation of 16 phage recombinants, yet failed to produce an overlapping series of sequences linking all the β -HCG genes. In fact, the isolates in the two groups *a* and *b* (Fig. 1) are remarkably similar, with clustered end points. It is possible that as there are repeats and symmetries in this locus, certain regions are unstable when cloned in *Escherichia coli*. This explanation is supported by the fact that the coding sequences tend to be found close to the ends of the phage isolates. A random distribution of DNA around the hybridizing regions would normally be expected as there are 16 independently isolated phage recombinants from two separate libraries generated by different restriction enzymes. This positioning of coding sequences at the ends of the recombinants also makes it difficult to isolate a suitable fragment for use as a probe to 'walk' along the chromosome.

One potential region of overlap between the sequences shown in Fig. 1 should be considered; it is possible that β -HCG 4 and 5 are identical and provide an overlap of ~1.5 kb between groups *a* and *b*. Some evidence in support of this overlap comes from the filter hybridizations of genomic DNA. With *BamHI*, both probes (Figs 4, 5) identify two bands, one of ~7 kb and the other >30 kb. Linkage of the two families as proposed above would generate a 7.2-kb *BamHI* fragment containing β -HCG 4/5 and 6. The other genes would be contained on

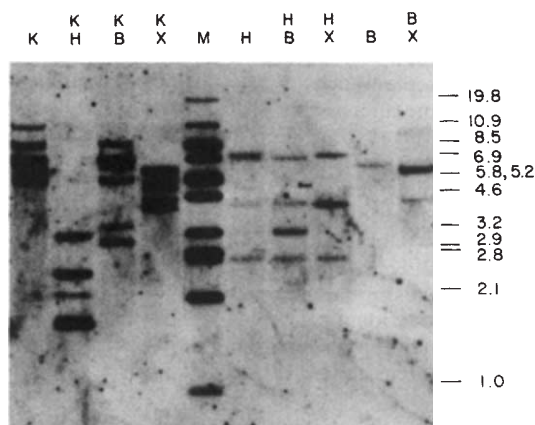


Fig. 4 Autoradiograph of a filter hybridization of human placental DNA; 10 μ g of DNA were digested with various restriction enzymes, electrophoresed through a 1.0% agarose gel and transferred to nitrocellulose sheets (Schleicher and Schuell) by the method of Southern⁷ as described previously⁴. The 32 P-labelled hybridization probe was the cloned β -HCG cDNA fragment (specific activity $1-2 \times 10^8$ c.p.m. μ g⁻¹; ref. 3) which is specific for the region of the gene(s) to the 3' side of the *HindIII* site (Fig. 3). K = *KpnI*, H = *HindIII*, B = *BamHI*, X = *XbaI*, M = markers whose sizes are given in kb at the right-hand side.

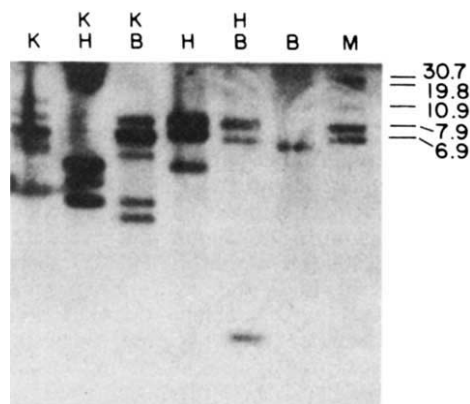


Fig. 5 Autoradiograph of a filter hybridization of human placental DNA digested with various restriction enzymes as in Fig. 4. The 32 P-labelled hybridization probe was a subcloned 0.8-kb *EcoRI*-*HindIII* fragment (Fig. 2). This probe is specific for the region of the gene(s) to the 5' side of the *HindIII* site (Fig. 3). K = *KpnI*, H = *HindIII*, B = *BamHI*, M = markers whose sizes are given in kb at the right-hand side.

the large *BamHI* fragment, which is consistent with the distribution of *BamHI* sites in the recombinants. However, in view of the fact that the β -HCG genes and surrounding regions are repeated, this is insufficient to establish an overlap, although it is suggestive.

Discussion

We have isolated eight separate genes that hybridize to the cDNA for the β -subunit of human chorionic gonadotropin. These genes have very similar structures and, at least in some cases, are linked. Although it is difficult to estimate the total number of β -HCG genes in the human chromosome, we believe that most of them have now been isolated. Until sequence analysis is complete, we cannot exclude the possibility that the eight genes include some pseudogenes or the related gene, β -LH (see below).

Four of the β -HCG genes (1, 2 and 5, 6) have been shown to exist as inverted pairs. In both cases the genes are arranged in the order 5'-3', space, 3'-5' so that, assuming both genes are active, transcription would be convergent. The 3' ends of the two genes are separated by ~2.25 kb. DNA sequencing and restriction enzyme analysis has shown that these two pairs are very similar. Inverted gene pairs have also been observed in the silk moth chorion genes, although in this case the pairs consist of dissimilar genes (type A and type B), transcription is divergent, and the distance between the 5' ends of the two members of the pair is only ~300 nucleotides⁸. In chicken, both divergent and convergent pairs of histone genes have been identified⁹. Two dissimilar genes, H2A and H4, are convergent while there are two identical copies of H3 which are divergent, with 5' ends separated by less than 1 kb. Also, in *Drosophila*, two different cuticle genes, II and III, form a closely linked divergent pair¹⁰.

The other four cloned β -HCG genes are arranged in two tandem pairs; β -HCG 3 and 4 and β -HCG 7 and 8. In both cases these genes are separated by ~5.5 kb. It is also possible to consider β -HCG 2 and 3 as a divergent pair of genes with 5' ends separated by ~8 kb. Examination of the distribution of restriction enzyme sites in Figs 1-3 reveals extensive regions of repeats: for example, the positions of the *KpnI*, *XbaI* and *XhoI* sites preceding β -HCG 3 and β -HCG 8 are identical (Fig. 1a and c). The *KpnI* site is almost 5 kb upstream from the proposed transcriptional start site.

The significance of the complex organization of the β -HCG genes is not obvious. The inverted and tandem pairs presumably arose from a complicated series of gene duplications and inversions encompassing an area considerably larger than the coding regions themselves. During this process the locations and sizes of the two introns remained the same. As β -HCG is the most recently evolved of the four glycoprotein hormone β -subunits, and is considered to have a common ancestor with β -LH, to which it is most closely related¹, it will be interesting to investigate the organization of the β -LH genes and to compare them with β -HCG. We consider it unlikely that the β -LH genomic organization will show the same complex patterns because the duplications and inversions of the β -HCG sequences presumably arose after the separation of β -HCG from β -LH.

In the hybridization conditions used here, it is possible that the β -HCG hybridization probes could also cross-react with β -LH as there is 82% homology in amino acid sequence between the two β -subunits¹. Whether this cross-reactivity can be observed depends on the choice of nucleotides in the third positions of the β -LH codons. The homology of β -HCG with β -FSH or β -TSH is too low to expect any cross-hybridization to take place¹. From DNA sequence and restriction enzyme analysis, we have found no evidence of β -LH gene sequences among the cloned fragments. This does not exclude the possibility, however, that β -LH is so closely related to β -HCG that one of the genes (β -HCG 1, 4, 7 and 8) for which we have no DNA sequence, could encode β -LH. It is also possible that a more faintly hybridizing band of total DNA, such as the 2.2-kb *HindIII*-*KpnI* fragment (Fig. 4), which has not yet been cloned, contained β -LH sequences.

The finding that the β -subunit of HCG is coded for by several genes while the α -subunit, which is common to all four glycoprotein hormones, seems to be encoded by a single gene^{4,5}, is surprising. In the placenta there is an excess of free α -subunit polypeptide over β -subunit polypeptide¹¹, thus it might be predicted that there would be multiple α - rather than β -genes. It will be interesting to establish whether all the β -subunit genes are functional and, if so, whether there are any significant differences between them and between their products.

This work was supported by a grant from the NIH. We thank T. Maniatis and E. Fritsch for the phage libraries, D. Hanahan for the DHI transformation protocol and E. Stavnezer for pSM804; P. Barkley, M. Goodwin and M. Ockler for help with the manuscript and art work; T. Grodzicker, D. Hanahan, S. Hughes, J. Sambrook, J. Sorge and K. Talmadge for valuable advice and discussions; and M. DeLuca for technical help in analysing one of the phage λ isolates.

Note added in proof: We have now established that the gene identified as β -HCG 4 is in fact β -LH (K. Talmadge and J.C.F.). This precludes the possibility that groups *a* and *b* overlap.

Received 5 August; accepted 11 October 1982.

- Pierce, J. G. & Parsons, T. F. *Rev. Biochem.* **50**, 465-495 (1981).
- Fiddes, J. C. & Goodman, H. M. *Nature* **281**, 351-356 (1979).
- Fiddes, J. C. & Goodman, H. M. *Nature* **286**, 684-687 (1980).
- Fiddes, J. C. & Goodman, H. M. *J. molec. appl. Genet.* **1**, 3-18 (1981).
- Boothby, M., Ruddon, R. W., Anderson, C., McWilliams, D. & Boime, I. *J. biol. Chem.* **255**, 5121-5127 (1981).
- Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157-1174 (1981).

- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Jones, C. W. & Kafatos, F. C. *Nature* **284**, 635-637 (1980); *Cell* **22**, 855-867 (1980).
- Harvey, R. P., Krieg, P. A., Robins, J., Coles, L. S. & Wells, J. R. E. *Nature* **294**, 49-53 (1981).
- Snyder, M. *et al. Cell* **29**, 1027-1040 (1982).
- Franchimont, P., Gaspard, V., Reuter, A. & Heynen, G. *Clin. Endocr.* **1**, 315-316 (1973).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237-251 (1977).
- Taylor, J. M., Illemensee, R. & Summers, S. *Biochim. biophys. Acta* **442**, 324-330 (1976).
- Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
- Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

LETTERS TO NATURE

Diameters and albedos of satellites of Uranus

R. Hamilton Brown

Department of Physics and Astronomy, and Planetary Geosciences Division, Hawaii Institute of Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA

Dale P. Cruikshank & David Morrison

Institute for Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA

The five known satellites of Uranus are too small and distant to permit direct measurement of their diameters. Consequently, products of the masses of the satellites, and estimates of their bulk densities and surface albedos have been used in discussions of their probable dimensions in the absence of measurements. The presence of water ice on the surfaces of Ariel, Umbriel, Titania and Oberon has been established spectrophotometrically¹⁻⁴ and the brightnesses of the satellites have been measured photoelectrically⁵. Determination of the masses of the uranian satellites depends on observations of perturbations of the satellite orbits and results chiefly in upper limits to the products of the masses of various pairs of satellites^{6,7}. The new diameter measurements reported here were made using the photometric/radiometric technique of diameter determination. This widely used technique has enabled the measurement of the diameters and albedos of approximately 250 asteroids and many planetary satellites. A recent recalibration of the technique using independent measurements of the diameters of three Solar System objects results in absolute accuracies of better than 5%. Our new albedo measurements show that Umbriel, Titania and Oberon are similar in albedo to J4 Callisto while Ariel is very similar in albedo to S7 Hyperion; the diameters of all four satellites are similar to those of the large, icy saturnian satellites S5 Rhea and S8 Iapetus.

The observations reported here were made with the 3.0-m Infrared Telescope Facility (IRTF) on Mauna Kea, Hawaii. The data were obtained on three nights in May 1982, as noted in Table 1. During this period, similar observations were made of Pluto and Neptune's satellite Triton without detections of their thermal radiation; results on these objects are given in the accompanying paper⁸.

A standard broadband *Q* filter which transmits from 16 to 26 μm was used for all observations. The star α Boo, for which the *Q* magnitude was taken to be -3.32 , was the primary standard⁹. Extinction on all nights averaged about 0.30 mag per airmass, which is normal for Mauna Kea; the volcanic haze covering part of the Earth's Northern Hemisphere produced no noticeable change in the extinction at 20 μm . Observations of Saturn's satellite S5 Rhea, whose radiated thermal energy

distribution is similar to that of the uranian satellites, and whose diameter ($1,530 \pm 10$ km) and photometric properties are well known from Voyager flybys^{10,11}, provided a consistency check on the calculated diameters of the uranian satellites. To ensure complete cancellation of the scattered thermal radiation from Uranus, the oscillating secondary mirror of the 3-m telescope was adjusted in amplitude and position angle of the oscillation so as to be perpendicular to the line joining a particular satellite and the planet. A 25-min integration on the sky diametrically opposite to the position of Ariel with respect to Uranus, where no satellite was located, resulted in a zero signal, confirming the cancellation of the flux contribution from the scattered thermal radiation of Uranus.

The calculation of diameters and albedos requires that monochromatic IR fluxes be determined from broadband observations. Because the colour temperature of the thermal radiation from airless bodies at the distance of the uranian satellites is quite low (≈ 80 K) and very different from that of a typical 20- μm stellar standard ($\approx 4,000$ K), close attention must be paid to the proper correction of the observed 16–26 μm flux ratios to monochromatic flux ratios at 20 μm . These corrections were made using determinations of the water vapour content of the atmosphere above the observatory derived from near IR spectra obtained on the same or adjacent nights (calibrated against a laboratory cell¹²), as well as measurements of the local atmospheric relative humidity throughout the nights made by the Kitt Peak National Observatory sky survey team. Transmission spectra of Mauna Kea model atmospheres, computed by V. Kunde for various values of precipitable water, were convolved with our filter transmission properties to derive the site-and-instrument-specific correction factors; calculated for each night because of the variability of atmospheric water vapour evidenced in the contemporaneous near-IR spectra. A detailed discussion of the monochromatic correction technique is given by Morrison¹³. The specific correction factors are listed in Table 1.

In addition to the IR observations, we have used new values of the brightnesses of the satellites in the *V* passband (central wavelength 0.55 μm) obtained 3 or 4 days after the IR measurements and at essentially the same solar phase angle. This angle, 0.03° , is very close to the minimum observable phase angle with the consequence that the visual magnitudes derived are some 0.4 mag smaller than published photometry⁵; in 1981 a strong, near-IR-brightness surge at opposition was discovered³. Since the uranian system is now almost pole-on as seen from Earth, the satellites, which are in circular orbits in Uranus' equatorial plane, are not expected to show brightness variations due to orbital motion. A detailed analysis of the opposition brightness surge will be published elsewhere¹⁴.

For calculation of the diameters and albedos of the uranian satellites we use the standard radiometric model¹³ and follow exactly the formulation used by Brown *et al.*¹⁵. The method is based on the determination of both reflected and absorbed

Table 1 Observations

Object	1982 dates	Integration time (min)	<i>Q</i>	Δ (AU)	R (AU)	Φ (deg)	H ₂ O (mm/airmass)	Y
Ariel	22.50 May	40	4.35 ± 0.20	17.87	18.88	0	≈ 1.0	0.88
Umbriel	20.33 May	40	4.67 ± 0.15	17.87	18.88	0	2–3	1.10
Titania	20.50 May	30	3.96 ± 0.10	17.87	18.88	0	2–3	1.10
Oberon	19.40 May	50	3.93 ± 0.14	17.87	18.88	0	3–4	1.20
Rhea	17.25 May	25	1.27 ± 0.04	8.88	9.68	4	2–3	1.03
Rhea	22.40 May	25	1.19 ± 0.04	8.93	9.68	4	≈ 1.0	0.93

Q is the broadband 20- μm magnitude. Δ is the object–Earth distance. Φ is the solar phase angle. H₂O is the precipitable water vapour. *Y* is the monochromatic correction factor (to be multiplied by the observed flux to yield the monochromatic flux). All errors quoted are ± 1 s.d. of the mean.

Table 2 Calculated diameters and albedos

Object	m_{20}	$V(0^\circ)$	$V(1.0)$	q	Diameter (km)	p_v
Ariel	-1.78 ± 0.22	1.03 ± 0.05	1.3 ± 0.2	0.8 ± 0.1	$1,330 \pm 130$	0.30 ± 0.06
Umbriel	-1.70 ± 0.18	1.89 ± 0.05	2.2 ± 0.2	0.7 ± 0.1	$1,110 \pm 100$	0.19 ± 0.04
Titania	-2.41 ± 0.14	0.91 ± 0.05	1.2 ± 0.2	0.7 ± 0.1	$1,600 \pm 120$	0.23 ± 0.04
Oberon	-2.53 ± 0.17	1.10 ± 0.05	1.4 ± 0.2	0.7 ± 0.1	$1,630 \pm 140$	0.18 ± 0.04
Rhea	-3.54 ± 0.06	—	0.24	1.02	$1,530 \pm 45$	0.60 ± 0.04
Rhea	-3.53 ± 0.06	—	0.24	1.02	$1,530 \pm 45$	0.60 ± 0.04

m_{20} is the 20- μm monochromatic magnitude corrected to an Earth-object distance of 1 AU and 0° solar phase angle (assuming a brightening of $0.01 \text{ mag deg}^{-1}$). $V(0^\circ)$ is the visual magnitude at 0° solar phase angle corrected to 1 AU from the Earth and Sun (see Text). $V(1,0)$ is the $V(0^\circ)$ with the opposition surge removed (see text), q is the assumed phase integral (except the value for Rhea taken from ref. 11). p_v is the visual geometric albedo. All errors are ± 1 s.d. of the mean, except those quoted for the assumed phase integrals and the $V(1,0)$, which represent our estimates of a reasonable range of values.

solar radiation. For an airless, slowly rotating body, we assume that the surface is in approximate thermal equilibrium with the absorbed component of the incident energy. The absorbed power is reradiated and can be determined from a measurement of thermal IR emission, such as the 20- μm observations reported here, while the reflected power can be determined by conventional photometry.

With observations of reflected and reradiated solar energy, it is possible to obtain a solution for albedo and diameter that is only weakly dependent on assumed parameters, such as the detailed photometric or thermophysical properties of the observed objects. In the case of the uranian satellites, possible departures from thermal equilibrium due to rotation will be unimportant, as the satellites are nearly pole-on with respect to the Earth and Sun. As previously mentioned, a consistency check is provided by the observations of Rhea, an object of known diameter and photometric properties, with a surface temperature distribution similar to that of the uranian satellites. (For more complete descriptions of the method see refs 13, 16, 17, and 15 for details of the mathematical formulation and error analysis.)

The phase integrals of the uranian satellites cannot be measured from Earth and a plausible value must be assumed for each satellite to derive its diameter and albedo. Jupiter's satellite Callisto ($p_v = 0.19$) has a measured phase integral of 0.7 (ref. 18), which we consider appropriate for use in the reduction of the Umbriel, Oberon, and Titania data in view of their roughly equivalent albedos. As the albedo of Ariel is higher, a phase integral of 0.8 (a value measured for Jupiter's Ganymede¹⁸) was adopted. However, for relatively dark objects, the method of reduction to diameters and albedos is only weakly dependent on the choice of a phase integral. For the case of Ariel, a change of ± 0.1 in the assumed phase integral of 0.8, results in about a $\pm 3\%$ change in the derived diameter; for Umbriel, Titania, and Oberon, a change of ± 0.1 in their assumed phase integral of 0.7 results in a $\pm 2\%$ change in the derived diameters. The estimated uncertainties in the phase integral have been incorporated into the quoted uncertainties for the diameters and albedos of the four satellites.

As previously mentioned, the visual magnitudes $V(0^\circ)$ listed in Table 2 were obtained at a solar phase angle of 0.03° and the therefore contain a substantial opposition surge. The calibration of the radiometric diameter scale by Brown *et al.*¹⁵ uses visual magnitudes of the calibration objects which do not include the average opposition surge of $\approx 0.3 \text{ mag}$ (ref. 19). To ensure that the diameters of the uranian satellites are calculated with photometric parameters more closely analogous to those of the Brown *et al.* calibration objects, visual magnitudes 0.3 mag fainter than the $V(0^\circ)$ in Table 2 were used in the reduction to diameters. Since the resulting albedos of the uranian satellites are relatively low, the derived diameters will depend only weakly on the visual magnitude¹³. In contrast, the derived albedos are very sensitive to the visual magnitude and the large opposition surges of the uranian satellites imply that their albedos are a strong function of the viewing geometry. For the purpose of calculating the albedos seen at an arbitrary solar

phase angle, the satellite diameters should henceforth be considered independent of visual magnitude. The visual magnitudes used in the reduction to diameters are listed in Table 2 in the column $V(1,0)$. The error bars estimated for the diameters include the tabulated uncertainty in $V(1,0)$.

The diameters and albedos shown in Table 2 are the first measurements of these important quantities for any of the uranian satellites. Previous estimates of the diameters of the satellites of Uranus are based on, and very sensitive to, assumptions regarding the satellites' albedos, and, except those of a recent spectrophotometric study⁴, tended to underestimate the satellites' sizes (It is interesting that the newly measured albedos agree with those inferred from a proposed relation between albedo and *VJHK* colour indices among outer Solar System bodies²⁰). Ariel, Umbriel, Titania and Oberon all have diameters roughly similar to the large, icy, Saturn satellites, particularly S5 Rhea, but significantly lower albedos (Rhea's albedo is about 0.6) than all known saturnian satellites except Hyperion, Phoebe, and the dark side of Iapetus. Hyperion has an albedo of 0.3, roughly equal to the albedo of Ariel, and reflectance spectra of Ariel and Hyperion in the 1.5–2.5 μm region are strikingly similar^{3,21}. The relatively low albedo of Hyperion and its reddish appearance in Voyager images suggests the presence of a dark contaminant on/in its surface; Hyperion's spectral similarity to Ariel may indicate that the surface contaminants are similar in both bodies. This argument can be extended by analogy to Umbriel, Titania, and Oberon³.

A comparison with bodies in the jovian system shows that the albedos of the uranian satellites are intermediate to the albedos of Ganymede and Callisto, whose surfaces are known to be largely composed of water ice contaminated with darker particles²². The relatively low albedos of the uranian satellites suggest that a dark contaminant is present on/in their surfaces, though there may be significant differences in the degree of contamination, the chemistry of the contaminant, the details of intermixing and the microstate of the surfaces when compared with the jovian and saturnian satellites. The question of the nature of the contaminant(s) in the surfaces of the uranian satellites is being addressed by one of us (R.H.B.), and will be published elsewhere.

Table 3 Satellite masses

Object	Mass (10^{-5} Uranus)		
	$\bar{\rho} = 1.0 \text{ g cm}^{-3}$	$\bar{\rho} = 1.3$	$\bar{\rho} = 2.3$
Ariel	1.42	1.85	3.27
Umbriel	0.83	1.08	1.91
Titania	2.48	3.22	5.70
Oberon	2.62	3.41	6.03

The satellite mass is shown for different values of an assumed mean density. The mass units are 10^{-5} Uranus mass where the mass of Uranus is $8.66 \times 10^{25} \text{ kg}$. A density of 1 g cm^{-3} is approximately that of pure water, a density of 1.3 g cm^{-3} corresponds to $\sim 60\%$ water ice and 40% silicates by weight and 2.3 g cm^{-3} is typical of carbonaceous chondritic material.

Table 3 lists the masses of the uranian satellites assuming several values for the mean density of each satellite and using the radii from this study. From observations of orbital perturbations, Greenberg^{6,7} has derived an upper limit to the product of the masses of Ariel and Umbriel of $\leq 10^{-9} M^2$ (M = the mass of Uranus). If we assume that the mean densities ($\bar{\rho}$) of Ariel and Umbriel are approximately equal, and use the radii from this study, we find that the Greenberg upper limit is satisfied for $\bar{\rho} \leq 2.91 \text{ g cm}^{-3}$. Veillet and Rattier²³ have reported that new observations and analyses of Miranda's motion yield a value for the product of the masses of Ariel and Umbriel of $1.10 \pm 0.25 \times 10^{-10} M^2$ which, with our radii, is consistent only with mean densities of $\leq 1 \text{ g cm}^{-3}$ for both satellites. The same study yields $1.5 \times 10^{-4} M$ for the mass of Titania. This value with our radius yields a mean density of 6.0 g cm^{-3} , which we believe is unreasonably high for a satellite of mainly ice compo-

sition. Densities of $\approx 1.3 \text{ g cm}^{-3}$, typical of Saturn's icy satellites²⁴, and the densities derived from Greenberg's Ariel-Umbriel mass constraints are consistent with the observed icy character of the surfaces of the uranian satellites. Determination of the mean densities of the satellites is crucial to the understanding of the origin and subsequent evolution of the uranian system. With reliable measurements of the uranian satellite radii now in hand, observations from which the satellite masses can be derived become paramount.

The infrared Telescope Facility is operated by the University of Hawaii under contract to NASA. We thank NASA for support of this project through grants NSG 7323, NSG 7312, and NGL 12-001-057, Eric Becklin, R. W. Capps, David Greip, John Hamilton and all the staff at the IRTF for assistance and Jay Goguen for making the photometric results available in advance of publication.

Received 23 June; accepted 1 October 1982.

1. Cruikshank, D. P. *Icarus* **41**, 246-248 (1980).
2. Cruikshank, D. P. & Brown, R. H. *Icarus* **45**, 607-611 (1981).
3. Brown, R. H. & Cruikshank, D. P. *Icarus* (submitted).
4. Soiffer, B. T., Neugebauer, G. & Matthews, K. *Icarus* **45**, 612-617 (1981).
5. Cruikshank, D. P. in *Uranus and the Outer Planets* (ed. Hunt, G. E.) 193-210 (Cambridge University Press, 1982).
6. Greenberg, R. *Icarus* **24**, 325-332 (1975).
7. Greenberg, R. *Icarus* **29**, 427-433 (1976).
8. Morrison, D., Cruikshank, D. P. & Brown, R. H. *Nature* **300**, 425-427 (1982).
9. Simon, T., Morrison, D. & Cruikshank, D. P. *Astrophys. J.* **177**, L17-L20 (1972).
10. Smith, B. A. *et al. Science* **515**, 504-536 (1982).
11. Hanel, R. *et al. Science* **215**, 544-547 (1982).
12. Cruikshank, D. P. & Thompson, A. B. *Commun. Lunar Planet. Laboratory* **8**, 397-398 (1970).

13. Morrison, D. *Icarus* **19**, 1-14 (1973).
14. Brown, R. H. & Goguen, J. (in preparation) (1982).
15. Brown, R. H., Morrison, D., Tesco, C. M. & Brunk, W. E. *Icarus* (in the press).
16. Jones, T. J. & Morrison, D. *Astr. J.* **79**, 892-895 (1974).
17. Morrison, D. & Lebofsky, L. A. in *Asteroids* (ed. Gehrels, T.) 184-205 (University of Arizona Press, 1979).
18. Squyres, S. W. & Veverka, J. *Icarus* **46**, 137-155 (1981).
19. Bowell, E. & Lumme, K. in *Asteroids* (ed. Gehrels, T.) 119-167 (University of Arizona Press, 1979).
20. Hartmann, W. K., Cruikshank, D. P. & Degewij, J. *Icarus* (in the press).
21. Cruikshank, D. P. & Brown, R. H. *Icarus* **50**, 82-87 (1982).
22. Clark, R. N. *Icarus* **44**, 388-409 (1980).
23. Veillet, C. & Rattier, G. *Colloq. IAU* No. 60 (1981).
24. Tyler, G. L. *et al. Science* **215**, 553-557 (1982).

Diameters of Triton and Pluto

David Morrison & Dale P. Cruikshank

Institute for Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA

R. Hamilton Brown

Department of Physics and Astronomy and Planetary Geosciences Division, Hawaii Institute of Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA

Triton, the major satellite of Neptune, and Pluto, the outermost planet, are too small and distant to permit direct measurement of their size, and indirect arguments must be used to estimate diameter and albedo. These two objects invite a comparative study because they have essentially the same brightness (absolute visual magnitude $V(1, 0) \approx -1$) and current distance from the Sun (30 AU). In addition, the presence of methane frost on the surface, and of a possible methane atmosphere, have been detected for both objects from IR spectroscopy¹⁻⁴. On the basis of these similarities, Triton and Pluto were selected as targets for possible detection of thermal IR emission. We report here new upper limits to the thermal IR emission from these objects that permit us to set significant upper limits on their diameters and to demonstrate that both are high albedo objects. These results clearly exclude the possibility that Triton is the largest planetary satellite, and they are consistent with the small size of Pluto deduced from other data.

Our observations were made using a Ge bolometer at the 3.0-m Infrared Telescope Facility (IRTF) on Mauna Kea. This instrument, located at an elevation of 4.2 km, is among the most sensitive in the world for observations in the 20-40 μm spectral region where thermal radiation from these objects should be most easily detected. Observations were made on three nights (19, 20, and 22 May 1982) under photometrically clear skies. Atmospheric water vapour, as monitored from 1 to 3 μm spectrometry, declined over this observing period from about 3 mm precipitable H_2O to <1 mm. We were concerned

about the effect on the observations of the worldwide high-altitude volcanic haze, which was dramatically apparent near sunrise and sunset and produced anomalously high visible extinction, but the performance of the IRTF at 20 μm appeared to be unimpaired by this dust cloud.

All observations were made through a standard broad-band Q filter, covering wavelengths from ~ 16 to 26 μm . Stellar standards, and their adopted Q magnitudes, were α Lyr ($Q = -0.10$) and α Boo ($Q = -3.32$) (refs 5, 6). Measured stellar extinction coefficients were in the range 0.2-0.4 mag per air mass. As an additional standard with a spectral energy distribution more like that of Triton and Pluto, we selected Saturn's satellite S5 Rhea. Other programme objects observed on these nights were the uranian satellites Oberon, Titania, Umbriel, and Ariel; the results for these satellites together with additional details of the observing procedures are reported by Brown *et al.*⁷ in an accompanying paper.

Pluto was observed in excellent conditions on 20 May for a total of 4,200 s. Triton, which is at a more southerly declination and is therefore viewed through larger air masses, was observed for a total of 13,800 s, approximately equally distributed over three nights. Neither object was detected.

The observations are summarized in Table 1. The signal and noise levels are given in units of the signal through the Q filter from a star of magnitude $Q = 0$, as determined from the observations of α Boo and α Lyr. Each 'observation' consists of a 300-s integration period, made up of 15 pairs of 10-s integrations, with the programme object centred alternately in each beam. Each 300-s data block generates a value for the signal and for a standard deviation in the mean computed from the 15 individual integrations. These observations are then corrected for extinction and averaged to yield a single value for each night. The uncertainty is calculated both from the standard deviation of the 14-16 individual observations and from the average of the individual observations divided by $\sqrt{N}$, where N is the number of observations (14 or 16). These two estimates of uncertainty agreed well in all cases, demonstrating freedom from nonstatistical sources of error.

The final Q -band values for Triton and Pluto are given in Table 1. For Pluto we obtain $(3 \pm 20) \times 10^{-4}$ where the indicated uncertainty is the mean of the two values calculated. For Triton the final value is $(15 \pm 18) \times 10^{-4}$. In these units, a flux of

Table 1 IR observations 17–22 May 1982

Object	V(1, 0)	UT date	N*	Q-band signal†	$m_{20}‡$
Rhea	+0.24	17 May	5	0.31 ± 0.02	1.2
		22 May	5	0.34 ± 0.03	1.2
Pluto	-0.7	20 May	14	$(3 \pm 20) \times 10^{-4}$	7.1
Triton	-1.0	19 May	14	$(83 \pm 40) \times 10^{-4}$	7.2
		20 May	16	$(8 \pm 25) \times 10^{-4}$	
		22 May	16	$(-13 \pm 25) \times 10^{-4}$	

* Number of observations of 300 s each.

† Signal relative to that of a star with $Q = 0$.

‡ Monochromatic 20- μ m magnitude (corresponding to 1σ values for Pluto and Triton).

20×10^{-4} is equivalent to magnitude $Q = +6.8$. For both objects, the 3σ upper limit is $Q = +5.3$.

For a planetary object in approximate thermal equilibrium with the incident sunlight, the diameter and albedo can be calculated from two quantities: the absolute visual magnitude $V(1, 0)$ and the monochromatic 20- μ m magnitude m_{20} (refs 8, 9). We present results of these calculations for both the 1σ and 3σ values of Q given above.

The values of $V(1, 0)$ for Pluto and Triton are obtained from previous photometry and will be used here without correction for possible dependence on solar or orbital phase angles. For Pluto, we adopt the value of $B_0 = 15.98$ obtained in 1980 by Tedesco and Tholen¹⁰. For a colour $B - V = 0.79$ and a mean distance 39.4 AU, this value of B_0 implies $V(1, 0) = -0.7$. This absolute magnitude is fainter than that generally used in the past, reflecting the gradual decline in brightness of Pluto observed in recent years¹. For Triton, we use the value $V(1, 0) = -1.0$ obtained in 1972 and 1977 by Degewij *et al.*¹¹.

The Q magnitudes obtained in our observations are converted to monochromatic magnitudes (m_{20}) using the methods described by Morrison and others^{8,12}. We have calculated the sensitivity of our photometer as a function of wavelength over the Q filter band for a variety of atmospheric transmission functions, depending on the humidity. The corrections were made separately for each night, using the depth of the 1.4- μ m telluric water band and the value of the Q -magnitude extinction coefficient as indicators of the absorption by telluric H_2O in the Q band. The calculated values of m_{20} are given in the final column of Table 1. The 1σ and 3σ values of m_{20} adopted for our calculations are 7.1 and 5.6, corresponding to 20- μ m fluxes of 0.014 and 0.054 Jy, respectively.

The diameters and albedos were calculated from these magnitudes using the recent small revisions in the radiometric diameter scale by Brown *et al.*¹². We adopted for these calculations a phase integral $q = 1.0$, appropriate for objects with moderately high albedo. As a check, we used the same methodology to calculate the diameter of Rhea from observations made at the same time. For Rhea we obtain $R = 765 \pm 25$ km, in good agreement with the Voyager value of $R = 765 \pm 5$ km, for this satellite¹³.

For Triton, the 1σ limit yields $R = 1,800$ km and albedo 0.43, while the 3σ values are $R = 2,100$ km and 0.25. For Pluto, the 1σ limit yields $R = 1,500$ km and albedo 0.37, the 3σ values are $R = 1,950$ km and 0.22.

The diameter of Triton has long been a subject of speculation but of few firm data. Because its absolute brightness $[V(1, 0)]$ is comparable to that of Titan and Callisto, it has frequently been noted that Triton could, in fact, be the largest of the satellites, if its albedo were low (<0.1). In 1954 Kuiper attempted to measure its size directly with the 5-m Hale telescope, estimating $R = 1,900$ km (ref. 14), but the uncertainty in this value was both large and poorly defined¹⁵. In reviews published in 1974 and 1977^{16,17} only upper and lower limits to the radius of 6,500 and 1,100 km, respectively, were listed based on plausible albedos.

In 1978 Cruikshank *et al.*¹⁵ presented spectroscopic evidence that the surface of Triton was 'rocky', leading them to estimate

that the albedo was confined to the range $0.02 \leq p_v \leq 0.5$, corresponding to radii between 7,400 and 1,500 km. These authors also used the radiometric technique to set a 3σ limit of $m_{20} > 4.3$, constraining the radius to be smaller than $\sim 2,600$ km, which is very near to the radius of the largest satellites: Ganymede, Callisto, and Titan. Thus their final estimate for the radius of Triton was $2,600 \geq R \geq 1,500$ km.

The 3σ upper limit for the radius of Triton set in this study is 2,100 km, 80% of the limiting value obtained by Cruikshank *et al.*¹⁵. To estimate a lower limit, we reason by analogy with other outer Solar System satellites. With but one exception, the icy satellites of Jupiter and Saturn have albedos ≤ 0.7 . (Enceladus, with $p_v \approx 1.0$, is an exceptional object, apparently subject to a long history of surface activity)¹⁸. An additional constraint is suggested by the downturn in Triton's spectral reflectivity in the UV, which resembles that of Europa ($p_v \approx 0.6$). We adopt here a limiting albedo $p_v < 0.7$, corresponding to $R > 1,300$ km.

The range of sizes for Triton quoted above ($2,100 > R > 1,300$ km) represents limits, not standard errors in the usual statistical sense. Our conclusions about the size and albedo of Triton, presented in the more conventional form of a probable value and standard deviation, would be $R = 1,600 \pm 200$ km, $p_v = 0.4 \pm 0.1$. This is the same size range as Io and Europa, and it is clearly smaller than the giant satellites Ganymede, Callisto, and Titan. The albedo is that of an icy surface with no more than moderate contamination, roughly similar to Ganymede in the jovian system or to the major icy satellites of Saturn, but distinctly brighter than the uranian satellites.

The size of Pluto is unknown, although its mass is fairly well constrained by the known parameters of the orbit of Charon, Pluto's satellite. The first indication that Pluto might be a lunar-sized object ($R \sim 2,000$ km) was derived from the detection of methane ice on its surface, suggesting that it probably had a moderately high albedo¹. From its mass of 0.0022 Earth masses¹⁹ and the physical constraint that its density be unity or greater, a radius of $R < 1,460$ km is implied. However, recent speckle interferometric results^{20,21} give radii in the range 1,500–2,000 km. A radius in the middle or upper end of this range would lead to a density that appears implausibly low, but all the data are consistent with a radius $\sim 1,500$ km.

The 3σ upper limit to the size of Pluto derived from our observations is $R < 1,950$ km, while a more probable upper limit is $\sim 1,700$ km, consistent with the other estimates quoted above. The albedo of Pluto for $R = 1,500$ km and $V(1, 0) = -0.7$ is 0.4, similar to that judged most likely for Triton and rather typical of icy satellites in the jovian and saturnian systems, but not of the uranian satellites.

The radiometric upper limits derived in this paper for the diameters of Triton and Pluto are consistent with icy objects of sub-lunar size, and with the hypothesis that the two objects are physically similar to each other. If, as we suspect, they both have albedos ~ 0.4 , then about a factor of two improvement in IR systems will be required to measure their thermal emission.

We thank NASA for support of this project through grants NSG 7323, NSG 7312, and NGL 12-001-057, E. E. Becklin, R. W. Capps, D. Greip, J. Hamilton, and the IRTF staff for their special support of the instrumentation and observing at Mauna Kea.

Received 23 June; accepted 1 October 1982.

1. Cruikshank, D. P., Jones, T. J. & Pilcher, C. B. *Science* **194**, 835–837 (1976).
2. Cruikshank, D. P. & Silvestro, P. M. *Astrophys. J.* **233**, 1016–1020 (1979).
3. Soifer, B. T., Neugebauer, G. & Matthews, K. *Astr. J.* **85**, 166–167 (1980).
4. Fink, U. *et al. Icarus* **44**, 62–71 (1980).
5. Simon, T., Morrison, D. & Cruikshank, D. P. *Astrophys. J. Lett.* **177**, L17–L20 (1972).
6. Morrison, D. & Simon, T. *Astrophys. J.* **186**, 193–206 (1973).
7. Brown, R. H., Cruikshank, D. P. & Morrison, D. *Nature* **300**, 423–425 (1982).
8. Morrison, D. *Icarus* **19**, 1–14 (1973).
9. Jones, T. J. & Morrison, D. *Astr. J.* **79**, 892–895 (1974).
10. Tedesco, E. F. & Tholen, D. J. *Bull. Am. astr. Soc.* **12**, 729 (1980).
11. Degewij, J., Anderson, L. E. & Zellner, B. *Icarus* **44**, 520–540 (1980).
12. Brown, R. H., Morrison, D., Telesco, C. M. & Brunk, W. E. *Icarus* (in the press).
13. Smith, B. A. *et al. Science* **212**, 163–191 (1981).
14. Kuiper, G. P. *Trans. IAU* **9**, 250–251 (1957).

15. Cruikshank, D. P., Stockton, A., Dyck, H. M., Becklin, E. E. & Macy, W. Jr *Icarus* **40**, 104–114 (1979).
16. Morrison, D. & Cruikshank, D. P. *Space Sci. Rev.* **15**, 641–739 (1974).
17. Morrison, D., Cruikshank, D. P. & Burns, J. A. in *Planetary Satellites* (ed. Burns, J.) 3–17 (University of Arizona Press, 1977).
18. Smith, B. A. *et al. Science* **215**, 504–537 (1982).
19. Harrington, R. S. & Christy, J. W. *Astr. J.* **88**, 442–443 (1981).
20. Arnold, S. J., Bokseberg, A. & Sargent, W. L. W. *Astrophys. J. Lett.* **234**, L159–L163 (1979).
21. Bonneau, D. & Foy, R. *Astr. Astrophys.* **92**, L1–L4 (1980).

Flow and instability of a viscous current down a slope

Herbert E. Huppert

Department of Applied Mathematics and Theoretical Physics,
Silver Street, Cambridge CB3 9EW, UK

If viscous fluid is released on a horizontal surface it rapidly takes up a circular plan form as it spreads. This form is observed^{1,2} to be stable to any small disturbances which are initiated on the front due, for example, to irregularities in the horizontal surface or to chance perturbations. Alternatively, if some fluid is released onto a sloping surface—for example, some liquid detergent on a slanted plate—a quite different plan form occurs. One, two or more extended regions of fluid develop downslope, as shown in Fig. 1a, b. A situation intermediate between these two is now discussed. Consider a broad band of viscous fluid, uniform in depth across a slope, released so as to flow down a constant slope. By following the motion, which is initially independent of the cross-slope coordinate, the speed of advance and the depth of the flow before it breaks up into a series of waves of ever increasing amplitude can be determined. I present an expression for the wavelength of the front, which is determined by surface tension and is independent of the coefficient of viscosity.

With the coordinate system depicted in Fig. 2 and use of the approximations of lubrication theory³, the y-independent, down-slope momentum equation can be written as

$$0 = g \sin \alpha + \nu u_{zz} \quad (1)$$

In deriving equation (1) we have neglected both surface tension, the effects of which will be analysed below, and contact line effects⁴, which are negligible¹ under the assumption that the Bond number $B = \rho g l^2 / T \gg 1$, where ρ is the fluid density, l a representative length scale of the current and T the surface tension. Use of the equation of continuity then leads to

$$h_t + (g \sin \alpha / \nu) h^2 h_x = 0 \quad (2)$$

as the nonlinear partial differential equation for the unknown free surface $h(x, t)$. To equation (2) must be added the global continuity equation

$$\int_0^{x_N(t)} h(x, t) dx = A \quad (3)$$

where $x_N(t)$ is the value of x at the front of the current and A is the initial cross-sectional area. The particular solution of equations (2) and (3) is sought in the range $0 \leq x \leq x_N(t)$.

Equation (2) shows that h is constant along characteristics given by

$$\frac{dx}{dt} = (g \sin \alpha / \nu) h^2 \quad (4)$$

Thus if initially $h = f(x)$, say, the equation of the characteristics is given by

$$x = x_0 + (g \sin \alpha / \nu) f^2(x_0) t \quad (5)$$

where x_0 is the initial value of the characteristic. The solution

of equation (2) is thus

$$h = [\nu(x - x_0) / g \sin \alpha]^{1/2} t^{-1/2} \quad (6a)$$

$$\rightarrow (\nu / g \sin \alpha)^{1/2} x^{1/2} t^{-1/2} \quad (x \gg x_0) \quad (6b)$$

independent of the initial conditions. When combined with equation (3), in order to evaluate the length of the current, equation (6b) proves that some time after the initiation of the current, no matter what the initial shape, the solution takes the form

$$h = (\nu / g \sin \alpha)^{1/2} x^{1/2} t^{-1/2} \quad (7)$$

$$0 \leq x \leq x_N = (9A^2 g \sin \alpha / 4\nu)^{1/3} t^{1/3}$$

Expressed alternatively, equation (7) represents the unique similarity solution of equations (2) and (3). The profile of the predicted current ends abruptly at $x = x_N$ with $h = h_N(t) = 1.5A/x_N$ there. The profile can be smoothed off at x_N by including the effects of surface tension. This will be done below after a discussion of an experimental investigation of the validity of the truncated profile expressed by equation (7).

Some 30 experiments were conducted using a pentagonal sheet of Perspex. The longest side was 101.7 cm and from it two sides of length 82 cm emanated at right angles. The remaining two sides were 60 cm. The sheet was surrounded by a Perspex wall 5 cm high to make a tray which was firmly attached to a rigid, flat board. Before each experiment the board was tilted by raising the 101.7-cm long side by the required amount. A removable Perspex gate was fitted 5.0 cm from the back edge of the tray and fluid poured into the space behind the gate. Three different fluids, whose physical properties are listed in Table 1, were used. After the fluid behind the gate had been left a sufficiently long time that it had become quite stationary, the gate was raised and the fluid proceeded to pour down the slope.

At first the motion was virtually independent of the cross-slope coordinate except near the walls where viscous drag retarded the flow. This form of motion is clearly evident in Fig. 1c. Observations of the length of the current as a function of time were taken and some typical results plotted in Fig. 3. It can be seen that the agreement between the experimental observations and the theoretical prediction (7) is good. The agreement justifies the neglect of both surface tension and contact line effects in predicting the temporal development of the two-dimensional current.

After the two-dimensional motion had continued for some time, the flow front seemed spontaneously to develop a series of small amplitude waves of fairly constant wavelength across the slope, as seen in Fig. 1d. The amplitude of the waves increased in time as the maxima (points furthest down the slope) travelled faster than the minima, as seen in Fig. 1e. The wavelength remained unaltered. The long-time shape of the flow front was different for the two fluids used. Both silicone oils developed the form shown in Fig. 1f. This consisted of a periodic, triangular front with tightly rounded maxima. The maxima were connected by very straight portions at an angle to the slope to extremely pointed minima. The glycerine developed the form shown in Fig. 1g. This was also periodic, though with much less tightly rounded maxima. These were again connected by extremely straight portions, but in this case they were almost directly down slope and connected to very broad minima.

Table 1 The fluids used, their viscosity and surface tension

Fluid	Viscosity at 17 °C (cm ² s ⁻¹)	Surface tension at 17 °C (dyn cm ⁻¹)
Silicone oil MS200/100	1.15	20.6
Silicone oil MS200/1000	12.8	22.0
Glycerine	9.8	60.2

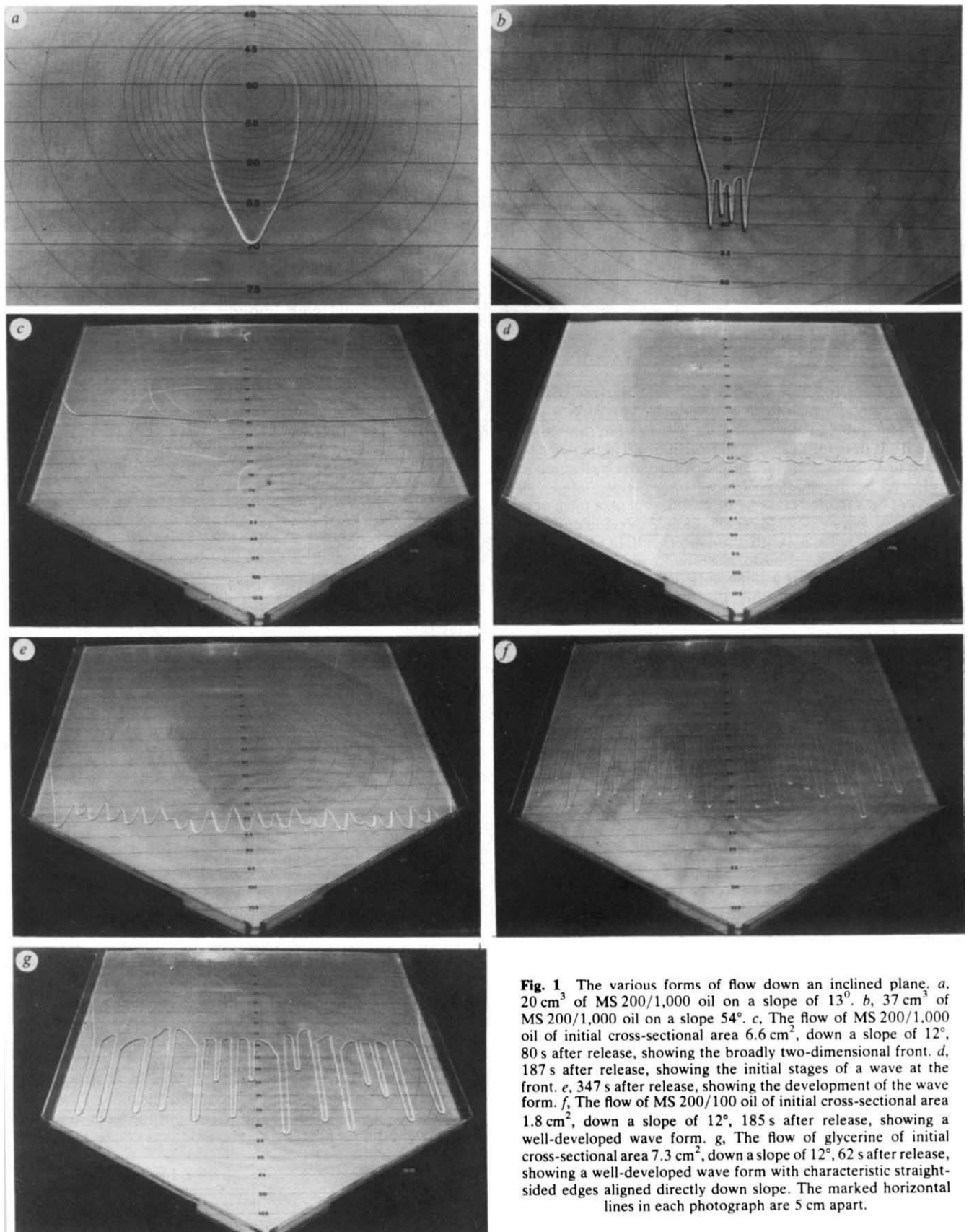


Fig. 1 The various forms of flow down an inclined plane. *a*, 20 cm^3 of MS 200/1,000 oil on a slope of 13° . *b*, 37 cm^3 of MS 200/1,000 oil on a slope 54° . *c*, The flow of MS 200/1,000 oil of initial cross-sectional area 6.6 cm^2 , down a slope of 12° , 80 s after release, showing the broadly two-dimensional front. *d*, 187 s after release, showing the initial stages of a wave at the front. *e*, 347 s after release, showing the development of the wave form. *f*, The flow of MS 200/100 oil of initial cross-sectional area 1.8 cm^2 , down a slope of 12° , 185 s after release, showing a well-developed wave form. *g*, The flow of glycerine of initial cross-sectional area 7.3 cm^2 , down a slope of 12° , 62 s after release, showing a well-developed wave form with characteristic straight-sided edges aligned directly down slope. The marked horizontal lines in each photograph are 5 cm apart.

The instability cannot be explained by viscous effects alone, since any cross-slope variation in depth of the current tends to be diminished, that is stabilized, by the cross-slope pressure gradient the variation induces. Surface tension must be taken into account.

Experiments with silicone oils whose coefficients of viscosity differ by an order of magnitude, although their surface tensions are almost equal, indicated that the wavelength of the instability is independent of the coefficient of viscosity. Experiments with the two fluids of comparable viscosity yet different surface

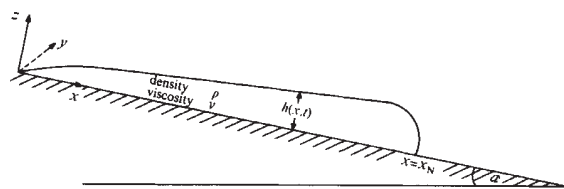


Fig. 2 A sketch of the flow and coordinate system.

tensions indicated that the wavelength is a function of the surface tension. Finally, observations taken during experiments using the same fluid but with amounts differing by up to a factor of 10 indicated that the wavelength is weakly dependent on the initial cross-sectional area A .

The form of the quasi-steady, two-dimensional tip, which is determined by including surface tension and by matching the tip onto the main flow given by equation (7), can be easily determined. The addition to equation (2) of the terms due to surface tension T leads to

$$h_t + (g \sin \alpha / \nu) h^2 h_x - 1/3 (T / \rho \nu) h^3 h_{xxx} = 0 \quad (8)$$

In the tip the dominant balance is between the second and third terms of equation (8), the solution to which can be written as

$$h = h_N(t) H(\xi) \quad \xi = (\rho g \sin \alpha / Th_N)^{1/3} (x_N - x) \quad (9)$$

where $H(\xi)$ satisfies

$$H^3 H''' + H^3 = 1 \quad (10)$$

$$H \rightarrow \left(\frac{16}{15}\right)^{1/4} \xi^{3/4} \quad (\xi \rightarrow 0) \quad (11)$$

$$\rightarrow 1 \quad (\xi \rightarrow \infty) \quad (12)$$

The length scale of the tip is thus given by $(Th_N / \rho g \sin \alpha)^{1/3}$.

The experiments indicated that the instability occurs after the two-dimensional flow has propagated to a critical length X_N^c proportional to $A^{1/2}$. Evaluating the length-scale of the tip at this point and conjecturing that the wavelength, λ , is given, in scale, by this length, suggests that $\lambda \sim (A^{1/2} T / \rho g \sin \alpha)^{1/3}$. The results of the various observations, normalized in this way, are presented in Fig. 4. The data are seen to collapse together systematically with this normalization and are well represented by

$$\lambda = 7.5 (A^{1/2} T / \rho g \sin \alpha)^{1/3} \quad (13)$$

The wavelength remained constant as the amplitude of the instability increased. For the silicone oils, the distance from the origin of the furthest points down the slope increased with time like $t^{0.35}$ while the distance of the minima increased like $t^{0.28}$. For glycerine, the distance of the front of the instability, once well established, increased like $t^{0.6}$ while the minima were virtually stationary.

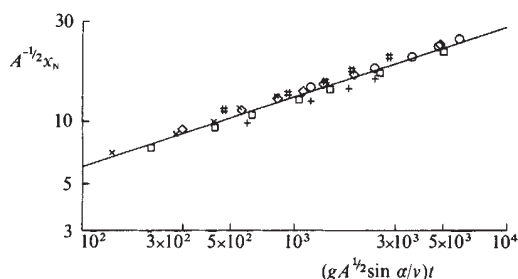


Fig. 3 The length of the two-dimensional flow, normalized with respect to $A^{1/2}$, as a function of a suitably non-dimensional time. The straight line is the theoretical prediction obtained from equation (7). The experimental points are for values of ν , $\sin \alpha$, A in cgs units of: $\square$, 12.9, 0.093, 8.9; $\diamond$, 12.9, 0.20, 3.1; $\circ$, 12.9, 0.65, 5.9; $+$, 9.8, 0.20, 9.0; $\times$, 9.8, 0.15, 3.6; $\#$, 1.2, 0.11, 1.0.

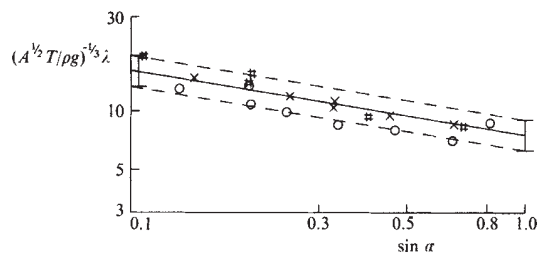


Fig. 4 The wavelength of the instability at the front, normalized with respect to $(A^{1/2} T / \rho g)^{1/3}$, as a function of the slope angle, α . Experiments with fluids of viscosities 12.9, 9.8 and $1.2 \text{ cm}^2 \text{ s}^{-1}$ are plotted as $\circ$, $\times$ and $\#$ respectively. The solid line is given by equation (14) and is the best-fit to the data. The two error bars at the end of the line represent $\pm 15\%$, a typical standard deviation of the wavelength across the slope. Where two or more data points fell on top of each other all but one were omitted for the sake of clarity.

I thank E. J. Hinch for stimulating and helpful conversations, many other of my colleagues in DAMTP for useful discussions, J. M. Wheeler for experimental help and for preparation of the figures, and D. C. Cheesley, D. Lipman, G. Parker and E. J. Sharp for technical assistance.

Received 14 June; accepted 8 September 1982.

1. Huppert, H. E. *J. Fluid Mech.* **121**, 43–58 (1982).
2. Huppert, H. E. *et al. J. Volcan. Geotherm. Res.* **14** (in the press).
3. Batchelor, G. K. *An Introduction to Fluid Dynamics* (Cambridge University Press, 1967).
4. Hocking, L. M. *Q. Jl Mech. appl. Math.* **34**, 37–55 (1981).

Forces between two adsorbed polyethylene oxide layers immersed in a good aqueous solvent

Jacob Klein*† & Paul Luckham*

* Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

† Polymer Department, Weizmann Institute, Rehovot, Israel

Recent measurements of the forces acting between two polystyrene layers adsorbed onto mica surfaces, immersed in a poor solvent for the polymer, show strong initial attraction as the surfaces approach, followed by ultimate repulsion¹. This has been attributed to attractive osmotic interactions between the adsorbed layers, and possibly to the effect of bridging^{2,3}. We have extended these measurements to the case of polyethylene oxide (PEO) layers adsorbed on mica, immersed in an aqueous 0.1 M KNO_3 medium at pH 6 (a good solvent for PEO). Using monodispersed polymer of two molecular weights, we find that an equilibrium force–distance profile is indicated. As the surfaces bearing the adsorbed PEO approach, repulsive forces commence at a surface separation $D \approx 6 \pm 1 R_g$ (unperturbed radius of gyration of the respective polymers) and increase monotonically on approach; on subsequent separation the forces decrease monotonically to zero. We find no evidence for attraction or adhesion between the adsorbed layers in this system.

The experimental technique and procedure closely resemble those previously described^{1,4}. The method permits the measurement of the force $F(D)$ acting between the mica surfaces a distance D apart, as well as the mean refractive index $n(D)$ of the medium separating them.

The forces were first determined between the bare mica surfaces in pure electrolyte. Polyethylene oxide solution was

then added to the required concentration (either 10^{-5} or $1.5 \times 10^{-4} \text{ g ml}^{-1}$), and the surfaces left to incubate for $16 \pm 2 \text{ h}$. Forces were then measured both in compression and decompression at different intervals (1 min to 1 h) and at various approach and separation rates (15–60 min per cycle), following which the PEO solution was replaced by pure solvent, and measurements repeated. The overall duration of experiments was generally between 48 and 72 h.

The PEO samples used were obtained from Toyo Soda (Japan). They were anionically prepared and their molecular characteristics (manufacturers data, determined by GPC) are shown below.

	M_w	M_w/M_n	R_g
PEO1	4.00×10^4	1.03	6.5 nm
PEO2	1.60×10^5	1.04	13 nm

The results are shown in Fig. 1. In the pure electrolyte (Fig. 1a) the forces between the mica surfaces are typical of electrostatic double-layer (DLVO) interactions⁴, with a secondary minimum as indicated in the inset and the characteristic strong short range repulsion (Fig. 1a). The surface potential of the mica deduced from the Debye length is $\sim 100 \text{ mV}$. The force-distance profiles as well as the surface potential correspond closely to those observed previously for mica surfaces in various aqueous electrolytes⁵.

Following adsorption of the polymer, the force profiles change as shown in Fig. 1a: no forces are observed as the surfaces approach from $D \approx 500 \text{ nm}$ down to $D \approx 6 \pm 1 R_g$ (corresponding to 35–40 nm for PEO1, 65–75 nm for PEO2), and thereafter monotonically increasing repulsion sets in. The surfaces were generally brought in to about $6 \pm 1 \text{ nm}$ separation. The shape of $F(D)$ on decompression and recompression of the surfaces depends to some extent on the rate at which these are carried out: if decompression is rapid (up to about 5 min for separation to $F(D) = 0$) the dashed curve (Fig. 1a) is followed. However, if separation is sufficiently slow (30–60 min for separation to $F(D) = 0$), the decompression curve merges (within error) into the original compression profile (solid curve). A similar effect is observed on recompression: if the surfaces are compressed, then rapidly separated, ($< 5 \text{ min}$) and then brought together again within 5 min, then the resulting $F(D)$ profile on subsequent compression follows the broken curve. If a longer time (30–60 min) elapses between separation and approach, the compression profile reverts to the original solid curve. Thus the variation of $F(D)$ with compression–decompression rates in these experiments is within the shaded regions indicated. We find no evidence for the build-up of thick adsorbed layers with time beyond the values corresponding to the solid curves in Fig. 1. The results described were obtained (within experimental error) for the two incubation concentrations used, both in solution and following the replacement of polymer solution by pure electrolyte. They suggest that there is a quasi-equilibrium force–distance profile for this system, represented by the solid curves, while the broken curves indicate the effect on $F(D)$ as a result of incomplete relaxation of the adsorbed polymeric layers at short times ($\approx 5 \text{ min}$).

Figure 1b shows the variation of refractive index $n(D)$ with surface separation D for PEO1. These values of $n(D)$ were measured both in compression and decompression, and were the same (within error) both in solution and following the replacement of solution by pure electrolyte, at the two incubation concentrations. They indicate (as for polystyrene adsorbed onto mica from cyclohexane) that there is little desorption or extrusion of polymers over the times and in the conditions of our experiments. The adsorbance Γ of polymer onto the mica may be estimated from $n(D)$ (ref. 1), and corresponds to $\Gamma \approx 5 \text{ mg m}^{-2}$ of mica surface. To our knowledge there are no independent measurements of PEO adsorbance onto plane surfaces to compare this value of Γ with, though we note that it is comparable to adsorbance values for other polymers on

plane surfaces⁵. It is, however, higher than PEO adsorbance onto colloidal silica particles⁶ (for which $\Gamma \sim 1 \text{ mg m}^{-2}$).

The onset of interactions between the adsorbed layers in the present study suggests an effective extension of polymer from each mica surface of some $3R_g$. This compares with values of $\sim 2R_g$ measured for various polymers (but not PEO) adsorbed on plane surface from θ -solvents⁷, and with an effective extension of some $1.5 R_g$ for polystyrene adsorbed onto mica in worse-than θ conditions¹. Our value of $2.5\text{--}3R_g$ is comparable with a hydrodynamic thickness of about $2R_g$ for PEO layers adsorbed onto polystyrene latex particles in a good aqueous solvent, as measured by light scattering⁸, and also with theoretically predicted values of some $3R_g$ for adsorbed poly-

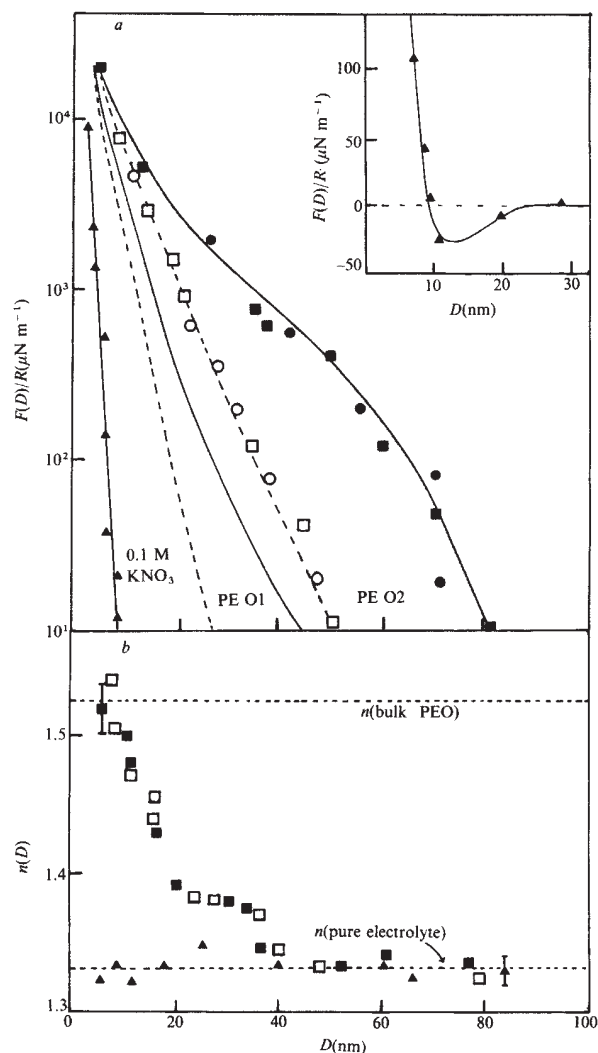


Fig. 1 a, Forces $F(D)$ between curved mica surface (mean radius of curvature $R \approx 0.5 \times 10^{-2} \text{ m}$) in 0.1 M KNO_3 (BDH Analar grade, in de-ionized double-distilled water) at $20 \pm 2^\circ \text{C}$. Δ , Force profile in pure electrolyte. The inset shows the secondary minimum on a linear scale. PEO2: Forces $F(D)$ following $16 \pm 2 \text{ h}$ immersion in $1.5 \times 10^{-4} \text{ g ml}^{-1}$ PEO2 solution. $\blacksquare$, Compression of surfaces following prolonged ($> 60 \text{ min}$) separation. $\square$, Compression $\sim 5 \text{ min}$ after rapid decompression of surfaces. $\bullet$, Slow decompression. $\circ$, Rapid decompression. Intermediate rates fall in shaded region. See text for details. PEO1: Forces $F(D)$ following $16 \pm 2 \text{ h}$ immersion in PEO1 solutions. Solid line, compression after prolonged ($> 60 \text{ min}$) separations; also slow decompression. Dashed line, rapid decompression. Both the pure electrolyte and the polymer solutions were filtered ($0.22 \mu\text{m}$ Millipore) before use. b, Mean refractive index $n(D)$ of medium separating mica surfaces. Δ , in 0.1 M KNO_3 ; $\blacksquare$, following $16 \pm 2 \text{ h}$ incubation in $10^{-5} \text{ g ml}^{-1}$ PEO1 solution; $\square$, following incubation in PEO1 solution and the subsequent replacement of the solution by pure electrolyte. Values unchanged over 24 h.

mers in an athermal solvent⁴. The monotonically repulsive nature of the interactions suggests that for this system there is little effect of bridging or desorption (which could lead to attraction)^{2,9}, the forces being due to an increasingly repulsive osmotic interaction as the opposing polymer layers come into overlap in the good solvent^{2,3,9}.

Our results contrast with the earlier study by Israelachvili and co-workers, who used a similar method to measure $F(D)$ between mica surfaces in an aqueous solution of highly polydisperse PEO¹⁰ (Carbowax resin WSR N80, Union Carbide). Their investigation indicated the continuous buildup with time of progressively thicker layers of surface adsorbed material (up to 1,000 nm over 24 h), as well as complex short- and long-time hysteretic behaviour suggestive of gel-like material on the surfaces. This may be due to the polydisperse nature of the polymer, as well as the possible presence of microgels or trace chemical impurities in the as-received Carbowax Resin (ref. 11 and J. N. Israelachvili, personal communication).

The present findings indicate that the interactions between adsorbed layers of monodispersed PEO across a good aqueous solvent are monotonically repulsive, and may be represented by a quasi-equilibrium force-distance profile at times longer than ~1 h. The extension of the adsorbed layers from each surface is somewhat larger than corresponding, independently measured values for polymers on plane surfaces in θ -solvents; the refractive index data indicates (in accordance with previous studies) a quasi-irreversible adsorption of polymer onto the mica surface.

We thank Professor D. Tabor for interest and helpful discussions, and Dr J. N. Israelachvili for useful suggestions.

Received 15 July; accepted 14 October 1982.

1. Klein, J. *Nature* **288**, 248–250 (1980); *JCS Faraday I* (in the press).
2. Scheutjens, J. M. H. M. & Fleer, G. J. *Adv. Colloid Interface Sci.* **16**, 361–381 (1982).
3. Klein, J. & Pincus, P. *Macromolecules* **15**, 1129–1135 (1982).
4. Israelachvili, J. N. & Adams, G. E. *JCS Faraday I* **74**, 975–1101 (1978).
5. Grant, W. H., Smith, L. E. & Stromberg, R. R. *Disc. Faraday Soc.* **59**, 209–217 (1976).
6. Rubio, J. & Kitchener, J. A. *J. Colloid Interface Sci.* **57**, 132–142 (1976).
7. Stromberg, R. R., Tutas, D. J. & Passaglia, E. *J. phys. Chem.* **69**, 3955–3964 (1965).
8. Barnett, K. G. *et al.* *Polymer* **22**, 283–284 (1981).
9. De Gennes, P. G. *Macromolecules* **15**, 492–500 (1982).
10. Israelachvili, J. N., Tandon, R. K. & White, L. R. *Nature* **277**, 120–121 (1979); *J. Colloid Interface Sci.* **78**, 432–446 (1980).
11. Union Carbide WSR-N80 Product Specification: 1-44K N80-1a, 70320 00, 19-6-1979.

Magnetostratigraphical dating of loess deposits in China

Friedrich Heller

Institut für Geophysik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

Liu Tung-sheng

Institute of Geology, Academia Sinica, PO Box 634, Beijing, China

The age of Chinese loess deposits has long been disputed. Biostratigraphical and earlier magnetostratigraphical investigations placed the entire loess formation within the Pleistocene. The new palaeomagnetic measurements reported here on a bore hole section near Lochuan (Shaanxi province) suggest a clearly defined magnetic polarity zonation which extends below the Olduvai subchron. A late Pliocene age of ~2.4 Myr is assigned to the oldest loess sediments measured. The intensity variations of natural remanent magnetization (NRM) and low field magnetic susceptibility are indicative of climatic changes during loess deposition.

In northern China nearly half a million square kilometres are covered by loess sediments. Along the middle Huangho valley these wind-transported, yellow to red-brown sediments occupy a coherent area of more than 300,000 km², known as the loess plateau¹. Here the loess deposits generally approach a thickness of 80–120 m. Typically they consist of an alternation

of (1) silty loess layers which developed in cold and arid climatic conditions and (2) clayey loess or palaeosoil beds which are indicative of warm and humid climate (see Fig. 1: succession of loess- and soil(S)-horizons). Some of the soil horizons are developed in such a way that they can be recognized and correlated over large distances. They serve as stratigraphical marker horizons². The environmental variations during loess formation are documented by faunal changes, especially of snail assemblages³, by the degree of crystallization of clay minerals⁴, and by carbonate content and other geochemical indicators⁵.

On the basis of lithology, unconformable contacts and erosion surfaces the loess deposit has been subdivided into three main stratigraphical units^{6–9} which all have been assigned a Pleistocene age: The Malan Loess of late Pleistocene age, The middle Pleistocene Lishih Loess and the early Pleistocene Wucheng Loess (Fig. 1). The biostratigraphical age estimates have been supported by several palaeomagnetic investigations^{10–12}. According to remanent magnetization directions obtained after alternating field (a.f.) demagnetization, the Brunhes–Matuyama boundary has been found to coincide roughly with the Wucheng–Lishih transition. The Jaramillo subchron was identified in the middle of the Wucheng division, thus the oldest strata were estimated to have an age of <1.2 Myr.

We have sampled a long core from a bore hole near Lochuan (Shaanxi province) ~800 km south-west of Beijing (lat. 35.4°N; long. 109.5°E) for palaeomagnetic analysis. The core penetrates a 136-m thick loess sequence which has been deposited on top of an Upper Pliocene (Hipparion bearing) lateritic substratum. More than 230 samples were obtained from the core with an average sampling interval of 60 cm. The samples were oriented relative to the vertical but not azimuthally. Before demagnetization the NRM intensity is relatively strong with a mean value of 5.08×10^{-2} Am⁻¹ for 150 soil samples, and 2.26×10^{-2} Am⁻¹ for 181 loess samples. The low field susceptibility of the soil horizons (mean: 2.49×10^{-3} SI units) also is about twice that of the loess layers (mean: 1.16×10^{-3} SI units). High field experiments show that the main magnetic minerals are magnetite and haematite. The NRM is strongly affected by secondary, viscous remanent magnetization (VRM) so that the NRM vector always gives inclinations consistent with the present field of the Earth. Advantage was taken of this property to detect several misoriented (inverted) samples. Because major portions of viscous magnetization (up to 80% of total NRM) are removed by low a.f.s of the order of 10–15 mT, the VRM apparently resides mainly in magnetite. However, viscosity tests in the laboratory show that progressive a.f. demagnetization increases the instability of NRM. A.f. cleaning of NRM often results in rather erratic or badly defined demagnetization curves. Thermal demagnetization also removes a large secondary VRM component up to 200 °C, but above 300 °C a well defined and stable component is observed which has either positive or negative inclination depending on the stratigraphical position. On orthogonal vector plots this characteristic component decreases unidirectionally towards the origin of projection on further heating. Maximum unblocking temperatures are usually above 575 °C. Thus the characteristic remanent magnetization of the loess is probably carried by haematite. Details of the loess rock magnetic properties will be described elsewhere.

After thermal demagnetization at 350 °C the core is subdivided into several clearly defined polarity zones, separated by sharp transitions (Fig. 1). In contrast to the results of earlier investigations which were based on a.f. cleaned NRM directions^{10–12}, the positive and negative inclinations average vectorially to the absolute value of the present local geomagnetic field inclination ($I = 52.5^\circ$). From top to bottom of the core the polarity changes are correlated with: (1) the Brunhes–Matuyama boundary at core depth 53.05 m; (2) the Jaramillo subchron between 67.30 and 72.50 m; (3) the Olduvai subchron between 107.40 and 113.10 m.

No definite evidence of shorter term geomagnetic excursions is indicated in our data. Three samples (out of a total of 231)

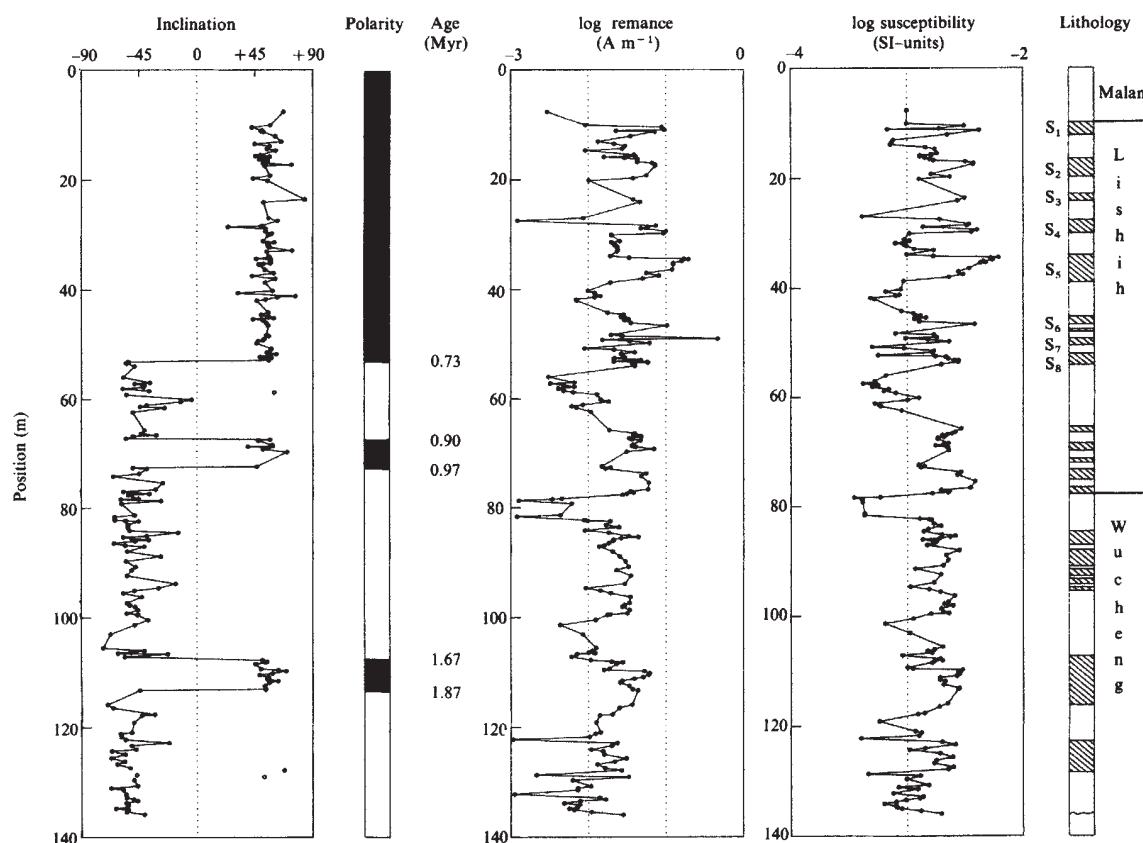


Fig. 1 Magnetic properties of the Lochuan bore hole versus core depth. Inclination data were mostly cleaned at 350 °C. The ages of polarity chron boundaries are according to Mankinen and Dalrymple¹³. Remanence intensities and susceptibilities are plotted before thermal cleaning. Lithological sequence of loess (white) and numbered soil layers (hatched) is according to Liu *et al.*².

do not fit to the observed polarity sequence. Possibly these samples may be completely remagnetized by a normal (present day) field or may suffer from unrecognized misorientation. Alternatively, they may indeed record field excursions or short polarity chrons (the two samples near 130 m depth would fit closely the Reunion events). Viscous remagnetization still affects the data to some extent, since the reversely magnetized portions of the core show higher directional scatter. No correlation is found between polarity and remanence intensity (Fig. 1) or low field susceptibility.

Our observations require a revision of some earlier ideas about the dating of Chinese loess. The Brunhes-Matuyama boundary is found to occur much higher in the loess profiles than thought previously¹⁰⁻¹². The Olduvai event has been identified for the first time, implying much greater ages for the entire loess formation. Converting depth into age by using the time scale of Mankinen and Dalrymple¹³ we notice an increase in apparent accumulation rate from 4.6 cm kyr⁻¹ below the Jaramillo event to 7.3 cm kyr⁻¹ in the upper 73 m. By linear extrapolation beyond the Olduvai event and due to the fact that the Matuyama-Gauss boundary is not observed in the loess sediments, we assign a minimum age of ~2.4 Myr to the beginning of loess sedimentation. This date coincides with age estimates for a general deterioration of climate in the Northern Hemisphere¹⁴. The age of the Plio-Pleistocene boundary is still controversial; however, if it were correlated with the top of the Olduvai event^{15,16}, then Chinese loess deposition, the older portions of which cover nearly the complete Matuyama epoch, started before the Pleistocene and dates back into the late Pliocene like the Central European loess deposits¹⁷. Also, the stratigraphical boundary between Lishih and Wucheng Loess no longer coincides with the Brunhes-Matuyama boundary, but has to be placed at ~1.1 Myr BP as it lies below the Jaramillo. Note that the negative polarity recorded in the bottom part of the loess profile (Fig. 1) continues down into the top of the Upper Pliocene red clay. After a few metres the

polarity in the clay reverses again to normal (not illustrated in Fig. 1). As the lithology changes from clay to loess soil rather gradually, it is tempting to assume continuous sedimentation between clay and loess so that this reversal would actually document the Matuyama-Gauss boundary.

The intensities of NRM (before cleaning) and low field susceptibility (Fig. 1) vary by 2-3 orders of magnitude throughout the section, being closely correlated to each other. High intensities characterize the soil horizons, whereas low values are typical of loess deposits¹¹. These features are ascribed to the pedogenetic processes of rock formation which favour haematite enrichment in the soils due to alteration in a relatively warm and humid climate. The carbonate content of the soil horizons is reduced² causing an overall volume decrease of up to 20%. Pre-existing magnetite becomes oxidized, the iron available from other minerals is mobilized and re-precipitates later in the form of oxides and hydroxides^{18,19}. Especially the latter process of formation of ferromagnetic minerals is thought to cause the intensity fluctuations of remanence and susceptibility which therefore are due to climatic changes and not to geomagnetic field variations as suggested by Wollin *et al.*²⁰ for Pleistocene deep sea sediments.

Thus the stable part of NRM may have a completely chemical origin. Using the magnetic intensity record we can estimate that about 17 glaciations have occurred in China since the Olduvai event. This number of glaciations is similar to that encountered in European loess deposits^{17,21}. It is also consistent with the number of oxygen isotope stages in post-Olduvai deep sea sediment from the Atlantic²². The Brunhes-Matuyama boundary is found in the soil layer S₈ (Fig. 1) which follows a loess unit deposited during a major glaciation period. Similar observations have been made in the oxygen isotope record of the deep sea sediments where this field polarity change is placed within the interglacial stage 19 (refs 22-24). The rock magnetic intensity variations become especially pronounced after 1.2 Myr ago (above core depth 82 m) and are interpreted to

record rather drastic climatic changes coinciding with major episodes of loess deposition at about 1.15, 0.85, 0.55, 0.44 and 0.18 Myr ago (age estimates by linear interpolation). The first of these episodes, presumably marking major glaciations, is accompanied by an increase of sedimentation rate by about a factor of two.

Our investigation demonstrates that, after thermal demagnetization, the NRM of the loess deposits in China renders a reliable record of geomagnetic field polarity. The rock magnetic properties apparently are largely controlled by chemical alteration processes during rock formation and therefore reflect climatic fluctuations during late Pliocene to Pleistocene times.

We thank Professors D. V. Kent, G. J. Kukla and W. Lowrie for critical reading of the manuscript. Ge T. M. and W. Gruber carried out many of the palaeomagnetic measurements. Contribution no. 387, Institut für Geophysik, ETH Zürich.

Received 26 July; accepted 5 October 1982.

1. Liu, T. S. & Chang, T. H. *Rep. 6th int. Congr. Quat. Geol.*, Warsaw 6, 503–524 (1964).
2. Liu, T. S., Wen, Q. Z., An, Z. S. & Zheng, H. H. *Int. Geol. Congr.*, Paris (1980).
3. Chen, D. N. *et al. Proc. 3rd natn. Quat. Congr.*, China (1979).
4. Han, J. M. *et al. Proc. 3rd natn. Quat. Congr.*, China (1979).
5. Wen, Q. Z. *et al. Proc. 3rd natn. Quat. Congr.*, China (1979).
6. Teilhard de Chardin, P. & Young, C. C. *Mem. geol. Surv. China* A, No. 8 (1930).
7. Liu, T. S. *Quat. Sinica* 1, 255–257 (1958).
8. Liu, T. S., Wang, T. M., Wang, K. L. & Wen, C. C. *Sci. Rec.* 2, 167–174 (1958).
9. Yen, C. *Quat. Sinica* 3, 88–103 (1960).
10. Li, H., An, Z. S. & Wang J. *Geochimica* 2, 93–103 (1974).
11. An, Z. S., Wang, J. & Li, H. *Geochimica* 4, 239–249 (1977).
12. Wang, R. Y., Wu, C. P., Ue, L. P., Chen, S. L. & Tuan, W. *Rev. Geol.* 26, 141–147 (1980).
13. Mankinen, E. A. & Dalrymple, G. B. *J. geophys. Res.* 84, 615–626 (1979).
14. Shackleton, N. J. & Opdyke, N. D. *Nature* 270, 216–219 (1977).
15. Van Couvering, J. A. *Geology* 6, 169 (1980).
16. Obradovich, J. D., Naeser, C. W., Izett, G. A., Pasini, G. & Bigazzi, G. *Nature* 298, 55–59 (1982).
17. Fink, J. & Kukla, G. J. *Quat. Res.* 7, 363–371 (1977).
18. Schwertmann, U. *Proc. Int. Clay Conf.*, Jerusalem 1, 159–165 (1966).
19. Tucholka, P. *Publ. Inst. Geophys. Pol. Acad. Sci. C-1(102)*, 127–141 (1976).
20. Wollin, G., Ericson, D. B. & Ryan, W. B. *Nature* 232, 549–551 (1971).
21. Kukla, G. J. in *After the Australopithecines* (eds Butzer, K. W. & Isaac, G. L.) 99–188 (Mouton, The Hague, 1975).
22. Van Donk, J. *Geol. Soc. Am. Mem.* 145, 147–163 (1976).
23. Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* 3, 39–55 (1973).
24. Shackleton, N. J. & Opdyke, N. D. *Geol. Soc. Am. Mem.* 145, 449–464 (1976).

Origin of desert loess from some experimental observations

W. B. Whalley*, J. R. Marshall† & B. J. Smith*

* Department of Geography, Queen's University of Belfast, Belfast BT7 1NN, UK

† NASA, Ames Research Center, Moffett Field, California 94035, USA

A longstanding problem of Quaternary geology has been the origin of the silty deposit known as loess. Of the many explanations for its formation and distribution^{1,2}, the most commonly accepted is that the silt is produced by subglacial grinding processes^{1,3,4}, followed by redistribution and deposition by the wind. The large expanses of loess in central Europe are therefore considered to be the consequence of Quaternary glaciation⁵. There is, however, a view^{6,7} that loessic silt may also originate as the product of processes in hot deserts. Some proponents of the desert loess hypothesis consider that silt-size material is the product of weathering processes⁷ which may include salt weathering⁸ or weathering down incipient cracks⁹. An alternative mechanism is aeolian attrition of sand grains, although, probably as a result of the experimental work of Kuenen¹⁰, Smalley and Krinsley¹¹ consider that only fine silt (2–6 μm) is produced in large quantities by the abrasion of quartz grains. We have simulated aeolian attrition of angular quartz grains previously produced by weathering in deserts. The products of abrasion show that both coarse and fine silt sizes are produced. These findings suggest that desert aeolian processes can produce loess. We also suggest that much of this material from many deserts has been deposited in the sea but that the Chinese loess could have been produced in the Gobi desert. The finest particles produced by such attrition may be a source of silica for silcrete formation.

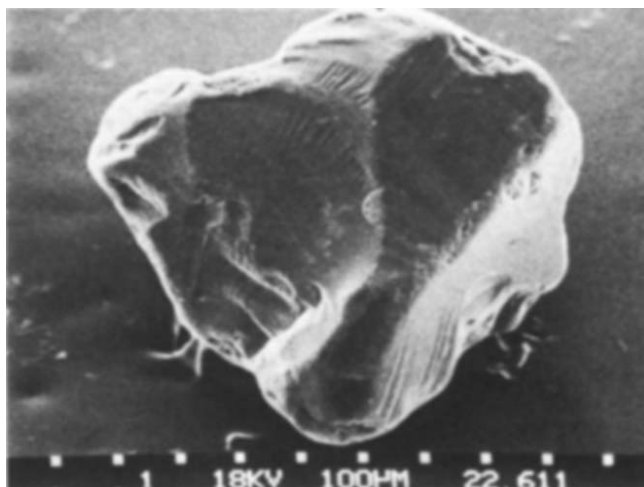


Fig. 1 An originally angular grain after 600 h of gentle erosion in the attrition device. At this stage the abrasion is confined to edges and corners but has already produced a considerable quantity of fragments. On this, and subsequent micrographs, the scale is given by the interval between the successive white squares at the bottom; the value (in μm) the second group of figures from the right.

We have concentrated on the aeolian attrition of quartz sand size grains released from the parent rock by weathering. In particular, whether there is a mechanism to produce particles in the 20–50 μm range from coarser material. Previous experiments on sand abrasion^{10,12} have tended to examine the larger products of attrition without either collecting or examining the smaller fragments chipped off the sand size grains. These fragments may, however, provide the very silt size material that the experiments were designed to detect. We have designed an experimental apparatus which, we believe, samples all the products of any attrition taking place. A triple-head 500-ml flask was used as an abrasion chamber in which ~10 g of crushed vein quartz (710–840 μm sieve size) was placed and agitated gently in a regulated airstream. The centre orifice was used to introduce a glass tube which vented air just above the base of the flask. The grains followed a regular saltating path with both interparticle collision and dune face avalanching. The 'air velocity' means little in these conditions but as field conditions are themselves very variable we do not consider this a limitation. Sampling of attrition debris was carried out in the outflow air at the top of two 300 mm high 'stilling wells' positioned on the flask side necks. This arrangement prevented the main, original particles from escaping. A scanning electron microscope (SEM) stub was placed in this airstream with double-stick tape on its surface. This was changed at intervals to allow periodic examination of small particles. The exhaust from the stilling wells was allowed to escape down polyethylene tubing. Fine particles are trapped by their own electrostatic charge in the first 10–20 mm and this region was removed at each sampling interval. The tube was cut and a portion examined by use of the SEM, together with fragments caught on the stub and 10 of the original grains in the flask. Collection efficiency is believed to be better than 95% of the fines produced. We have been able therefore to examine not only the abrading, original, grains and their progressive attrition but also the nature of the fine particles produced. Vein quartz, without apparent weaknesses introduced by the crushing, was used because the experiment was designed to look at the products of attrition as far as possible. Kuenen's experiment also used this type of material.

Figure 1 shows a grain after 600 h of gentle abrasion. Although it is rather poorly rounded compared with Fig. 2 the latter is from a desert dune environment¹³ and has been abraded for a very much longer period and probably at much higher tractive velocities. Indeed, it is surprising that such abrasion

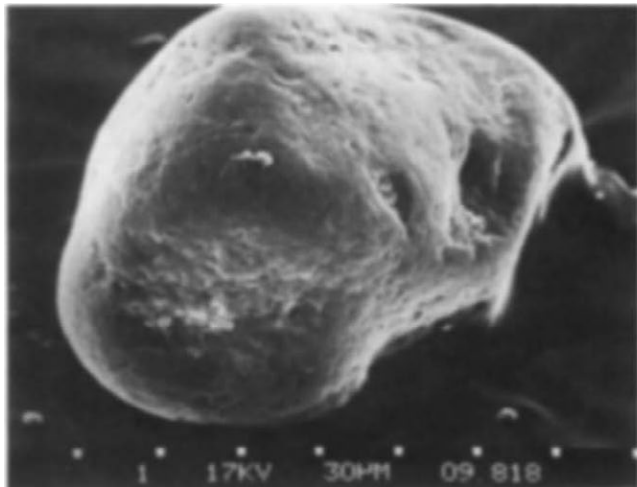


Fig. 2 This grain was taken from a fossil desert dune environment in Nigeria. There is considerably more rounding than seen in Fig. 1, abrasion now covers faces rather than just on corners. It is not known how long this grain might have been subjected to abrasive action but is likely to be very much longer than 600 h. The deposits are probably at least 25,000-yr old but the length of active dune formation is not known.

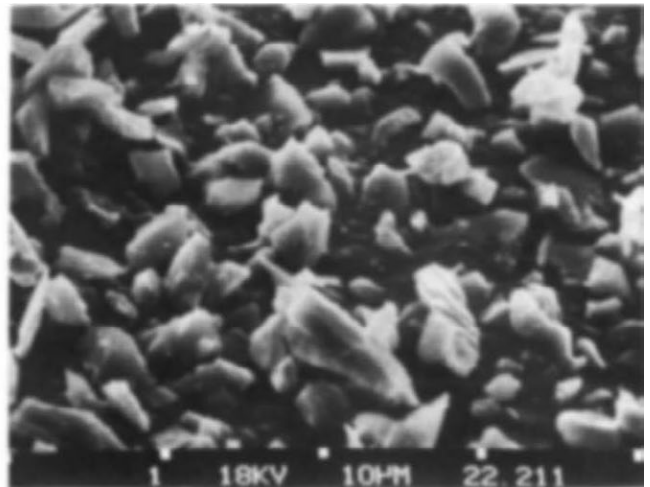


Fig. 4 Finer material than that seen in Fig. 3 is produced in the experiment. This micrograph shows fine silt and clay-size material caught in the polyethylene final trap.

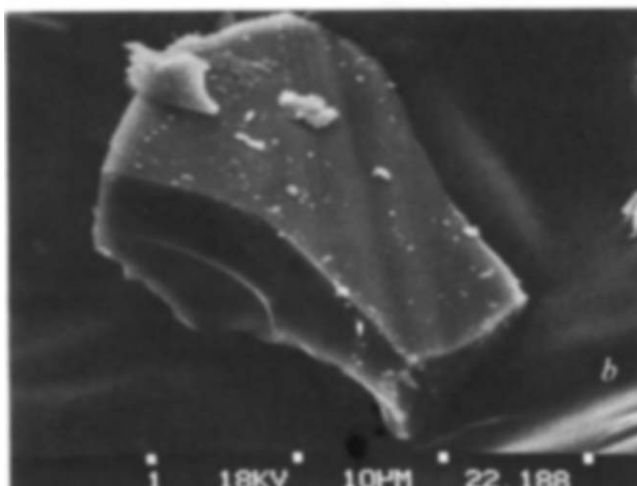
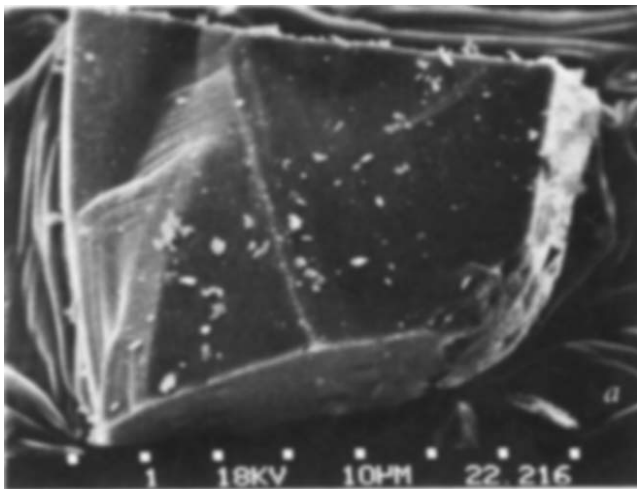


Fig. 3 Two coarse silt fragments (at roughly the extremes of this size range) produced by the attrition in the experiments reported. *a*, A grain with a section which has been rounded (between the arrows). The surface texture of this area is similar to that seen in Figs 1 and 2. *b*, A finer fragment with no such surface texture apparent. *a* was extracted after 200 h of attrition and *b* after 60 h.

seen in Fig. 1 has occurred. Not all desert grains are as rounded as that in Fig. 2 (ref. 14), this is possibly because they have not been abraded either for very long or sufficiently energetically. The abrasion process is one of progressive removal of small fragments, generally $<30\ \mu\text{m}$ but with rather fewer larger chips ($<100\ \mu\text{m}$). Figure 3 shows two of these larger fragments which, in this case, are products of edge rounding. Such rounding, however, is not often visible. Chips of this size may be produced in the manner suggested by Moss¹⁵ and Smalley and Vita-Finzi⁴ but are rather less common than these investigators imply. Kuenen's¹⁰ finding that "no quartz particles are produced in the size range dominating in loess" (although he did obtain some dust $<5\ \mu\text{m}$) may have been a major factor in the reluctance to accept aeolian abrasion as a means of producing loessic silt. Our experiments have shown that silt-size as well as some clay-size material is formed in low energy aeolian environments (Fig. 4). Exact determinations of the quantities of attrition products is not possible but (after 60 h) a mass reduction of about 1–5% was found; of this about 20% was $<5\ \mu\text{m}$. It is not yet known what are the relationships between particle velocity and size and quantity of fine material produced by aeolian action. In our experiment, fine material was still being produced after 600 h action. It is not known why Kuenen only detected the fine dust in his experiment. The surface textures seen in Figs 1 and 3a are very similar to that of the known aeolian origin of Fig. 2 and it seems reasonable to assume that the experimental production of fine material was matched naturally to give the rounded grain of Fig. 2.

There are several points of note. (1) Abrasion takes place primarily by point impacts of (originally) angular edges (Fig. 1) probably by grain spinning¹⁶. (2) Chips knocked off are of predominantly silt-size and are rather blocky in nature. This compares with published size ranges and SEM micrographs of loess^{17–19}. (3) Abrasion may be achieved at very low particle velocities. (4) There is some resemblance of rounded edges to so-called 'chemical precipitation' surface textures seen on SEM micrographs of quartz grains from desert dunes²⁰. (5) The very fine material produced in nature may be entrained by wind and removed completely from any desert system. The very finest abrasion products may even be dissolved and could be an important contributor to silicification of sediments²¹.

Finally, it seems that an aeolian abrasional origin for some loessic material is quite possible. It is unlikely that much European loess is produced in this way even though long distance travel from the Sahara is possible²². Most Saharan produced silt lies not on the southeastern Mediterranean land surface^{6,7} out in the Atlantic Ocean^{23,24} or even South America in the case of fine dust²⁵.

It would, for example, seem that a mechanism may exist whereby weathered grains in the Gobi desert could produce a substantial amount of China's loess. It is also possible that silts in the south-west of the United States and northern Mexico were produced by an aeolian abrasion mechanism, perhaps coupled with silt production by weathering mechanisms⁹.

We acknowledge the facilities provided by Professor D. H. Krinsley for the work described here and also to Queen's University Electron Microscope Unit.

Received 15 September; accepted 15 October 1982.

- Smalley, I. J. *Loess: Lithology and Genesis* (Dowden Hutchinson and Ross, New York, 1975).
- Lill, G. O. & Smalley, I. J. *Proc. geol. Ass.* **89**, 57–65 (1978).
- Smalley, I. J. *J. sedim. Petrol.* **36**, 669–676 (1966).
- Smalley, I. J. & Vita-Finzi, C. *J. sedim. Petrol.* **38**, 766–774 (1968).
- Smalley, I. J. & Leach, J. A. *Sedim. Geol.* **21**, 1–26 (1978).
- Yaalon, D. H. & Ganor, E. in *SCOPE Rep.* 14 (ed. Morales, C.) 187–193 (Wiley, New York, 1977).
- Yaalon, D. H. & Dan, J. Z. *Geomorph.* **20**, 91–105 (1974).
- Goudie, A. S., Cooke, R. U. & Doornkamp, J. C. J. *J. Arid Envir.* **2**, 105–112 (1979).
- Riezebos, P. A. & Van der Waals, L. *Sedim. Geol.* **12**, 279–285 (1974).
- Kuenen, Ph. H. *J. Geol.* **68**, 427–449 (1960).
- Smalley, I. J. & Krinsley, D. H. *Catena* **5**, 53–66 (1978).
- Linde, K. & Mycielska-Dowagialo, E. *Geogr. Annl.* **62A**, 171–184 (1980).
- Smith, B. J. & Whalley, W. B. *Catena* **8**, 345–367 (1981).
- Folk, R. L. *J. sedim. Petrol.* **52**, 93–101 (1978).
- Moss, A. J. *J. geol. Soc. Austr.* **13**, 97–136 (1966).
- Chepil, W. S. *Soil Sci.* **60**, 305–320 (1945).
- Smalley, I. J. & Cabrera, J. G. *Bull. geol. Soc. Am.* **81**, 1591–1596 (1970).
- Cegla, J., Buckley, T. & Smalley, I. J. *Sedimentology* **17**, 129–134 (1971).
- Whalley, W. B. *Geol. Mag.* **111**, 84–86 (1974).
- Marshall, J. R., Whalley, W. B. & Krinsley, D. H. *Geol. Soc. Am. Abstr. Proc.* 1000 (1981).
- Whalley, W. B. & Smith, B. J. *J. Arid Envir.* **4**, 21–29 (1981).
- Geovanoli, R. *Naturwissenschaften* **69**, 237–239 (1982).
- Glaccum, R. A. & Prospero, J. M. *Mar. Geol.* **37**, 295–321 (1980).
- Sarnthein, M. & Koopman, B. in *Sahara and Surrounding Seas* (eds Van Zinderen Bakker, Sr, E. M. & Coetzee, J. A.) 239–253 (Balkema, Rotterdam, 1980).
- Prospero, J. M., Glaccum, R. A. & Nees, R. T. *Nature* **289**, 570–572 (1981).

Trapped methyl radicals in chert

D. R. Griffiths, G. V. Robins & N. J. Seeley

University of London Institute of Archaeology,
31–34 Gordon Square, London WC1H 0PY, UK

H. Chandra, D. A. C. McNeil & M. C. R. Symons

Department of Chemistry, The University, Leicester LE1 7RH, UK

We have demonstrated recently that electron spin resonance (ESR) spectroscopy can detect a narrow signal from electron gain (E') centres in flint (Upper Cretaceous lepispheric chert)¹. This signal, caused by the effect of natural radiation, is lost on heating and may be useful for dating². ESR signals from various organic radicals are now reported in a wide variety of cherts. The thermal dependence of the signals may provide valuable information about the duration and temperature of ancient chert heatings.

The organic radical signals have been observed in Upper Cretaceous lepispheric cherts from Europe and the Middle East, and in Tertiary chalcidonic cherts from south-west France. The chalcidonic cherts comprise a coarse, fibrous α -quartz (chalcidony) and the lepispheric cherts predominantly comprise α -quartz lepispheres cemented by an interstitial micro-chalcidony.

The most important, and most unexpected result is the detection at ambient temperature of a well-defined quartet of lines (1:3:3:1) with a hyperfine coupling of 22.7G and a g -value of 2.0025 which is characteristic of methyl radicals³ (See Fig. 1 and Table 1). In addition a triplet due to two equivalent protons has been detected which we assign to $X\text{-}\dot{\text{C}}\text{H}_2$ radicals. Other unassigned features due to organic radicals have also been detected.

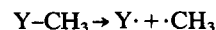
Most of the cherts studied have given a very weak methyl signal. After heating the chert, the signal is more intense, the methyl concentration reaching a maximum at around 350 °C before rapidly decreasing (Fig. 2). The temperature giving maximum methyl radical concentration varies with the duration of heating. Surprisingly, γ irradiation of chert did not alter the

Table 1 ESR data for methyl radicals in various environments, together with data for $X\text{-}\dot{\text{C}}\text{H}_2$ species

Radical	System temperature	g -value (average)	Hyperfine coupling (G)	Ref.
$\cdot\text{CH}_3$	$\text{CH}_3\text{I/vycor glass (RT)}$	—	22.7	4
$\cdot\text{CH}_3$	CH_4 (4K)	2.0025	22.9	5
$\cdot\text{CH}_3$	Chert (RT)	2.0025	22.7	This work
$\geq\text{SiOCH}_2$	$\text{CH}_3\text{OH/vycor glass (-140 °C)}$	2.0031	20.5	6
$X\text{-}\dot{\text{C}}\text{H}_2$	Chert (RT)	2.0031	21.0	This work

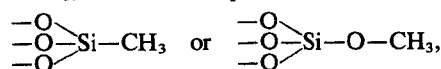
1G = 10^{-4} T; RT = room temperature

methyl radical concentration. γ irradiation followed by heating gave less growth of the methyl signal than did an identical heating of an unirradiated sample. It is thus necessary to postulate a radiation sensitive precursor ($Y\text{-}\dot{\text{C}}\text{H}_3$) which is thermally labile:



Several aspects of the environment of these $\cdot\text{CH}_3$ radicals can be delineated. The fact that the lines are narrow shows that the radicals are freely rotating, but not in the 'gas-phase'³. Also, the absence of 'spin-flip' satellite features at high power indicates that they are trapped in an environment free of paramagnetic nuclei³. Obviously, they cannot be close to any species with which they can react, and they are well separated from the radicals $Y\cdot$ (if they are radicals) because there is no spin-spin reaction broadening the lines. Almost certainly they are trapped in molecular-sized holes in a purely silica environment.

Similar comments apply to the $X\text{-}\dot{\text{C}}\text{H}_2$ radical. Furthermore, the group $X\cdot$ must contain only non-magnetic nuclei close to carbon as no further hyperfine coupling can be detected. γ irradiation increases the concentration of $X\text{-}\dot{\text{C}}\text{H}_2$ radicals. Their thermal behaviour is complex. The precursor of the methyl radicals, $Y\text{-}\dot{\text{C}}\text{H}_3$, is unknown. Species such as



though attractive, are unlikely as there is no reason to believe that the Si—C or O—C bonds are exceptionally weak. These bonds might be weakened by the proximity of a radical centre, but since the thermal yield of methyl is decreased by irradiation this is improbable. We favour the presence of some labile molecules trapped in the silica system, but cannot guess their identity at this stage. It is possible that the same species is the source of $X\text{-}\dot{\text{C}}\text{H}_2$ radicals on radiolysis (that is, $X \equiv Y$), as this

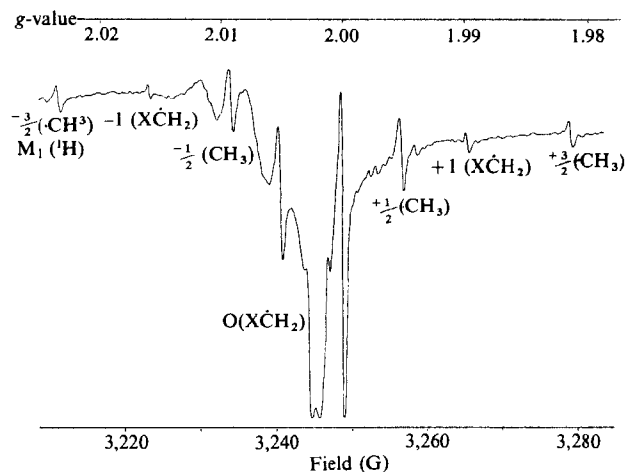


Fig. 1 Electron spin resonance spectrum of a chert sample from Ksar-Akil, Iraq, recorded at room temperature showing signals assigned to $\cdot\text{CH}_3$ and $X\text{-}\dot{\text{C}}\text{H}_2$ radicals. The sample volume was $\sim 0.2 \text{ cm}^3$. Spectra were recorded on a Varian E109 spectrometer capable of detecting $\sim 10^{11}$ spins.

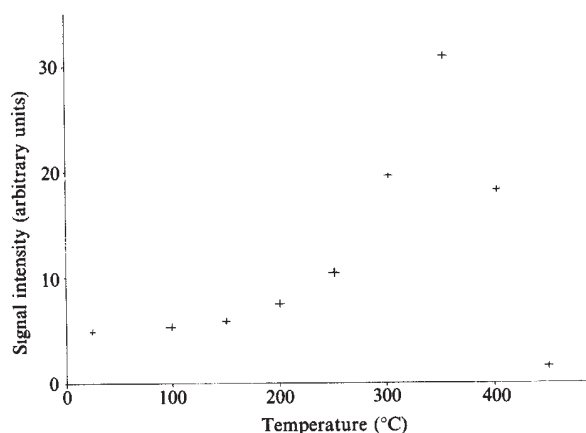


Fig. 2 Isochronal annealing curve showing the intensity of the methyl signal (measured at room temperature) after successive heating episodes of 15 min duration at each of the marked temperatures. The chert sample was from Al-Beidah, Jordan.

would account for the lower yield of methyl radicals on heating after irradiation. It seems to be important that many varieties of chert should contain a compound that is a thermal source of methyl radicals, since such compounds are most unusual.

The thermal behaviour of the organic radicals in chert is being further studied and is likely to prove useful in determining the temperature and duration of ancient heating episodes. Such information is of considerable interest in the study of ancient stone-working techniques. It is also useful in determining whether or not a sample has been sufficiently heated to be suitable for dating by thermoluminescence of ESR techniques.

We thank the SERC for financial assistance.

Received 2 July; accepted 23 September 1982.

1. Robins, G. V., Seeley, N. J., Symons, M. C. R. & McNeil, D. A. C. *Nature* **276**, 103 (1978).
2. Robins, G. V., Seeley, N. J., Symons, M. C. R. & Bowman, S. G. E. *Proc. 1980 Specialist Seminar on Thermoluminescence Dating*, Oxford (in the press).
3. Symons, M. C. R. *Chemical and Biochemical Aspects of Electron Spin Resonance Spectroscopy* (Van Nostrand Reinhold, London, 1978).
4. Turkevich, J. & Fujita, Y. *Science* **152**, 1619–1621 (1966).
5. Jen, C. K., Foner, S. N., Cochran, E. L. & Bowers, V. A. *Phys. Rev.* **112**, 1169–1182, (1958).
6. Melamud, E., Reisner, M. G. & Garbatski, U. *J. phys. Chem.* **77**, 1023–1027 (1973).

Effects of aerosols on photosynthesis

Siegfried A. W. Gerstl & Andrew Zardecki

Theoretical Division, University of California,
Los Alamos National Laboratory, T-DOT, MS-B279, Los Alamos,
New Mexico 87545, USA

Air pollution can cause various problems to agriculture through its direct or indirect effects on plants^{1,2}. The net effects of air pollution on agricultural crop yields are very difficult to estimate quantitatively, and are generally considered secondary to the effects of fertilizers or chemical disease control measures. However, continually increasing air pollution may represent a persistent and largely irreversible threat to agriculture in the future. Here we have examined only one indirect aerosol effect on agricultural plant life—the reduction in photosynthetically active radiation (PAR) reaching the ground in the presence of significant air pollution. This effect can be quantified confidently from results of detailed model calculations of the transfer of solar radiation through the atmosphere with varying levels of air pollution. For the biologically effective wavelength region between 0.35 and 0.72 μm , the reductions in PAR due to increasing tropospheric aerosol loads have been determined for rural and urban pollution scenarios. For example, for the north-eastern United States during the summer months, a reduction in PAR of about 20% and 33% was obtained when the visual range due to rural and urban aerosols, respectively, was decreased from clear air conditions to 10 km.

The effects of aerosols on the natural radiation balance in the atmosphere have been studied by using extensive measurement programmes³ and theoretical modelling⁴. Only recently, however, have detailed data on realistic air pollution scenarios become available to allow an accurate description of the optical parameters of aerosol-loaded atmospheres. For our model calculations we used the very detailed and recent data base for the optical properties of atmospheric aerosols compiled by the Air Force Geophysics Laboratory (AFGL)^{5,6}. Four altitude regimes with different aerosol models in each region are considered. Within the boundary layer (up to 2 km height), rural and urban-type aerosols are specified with varying meteorological ranges (surface visibilities). These models are intended to distinguish between aerosol types of natural and man-made origin over a land area. The 'rural model' represents aerosol conditions over continental areas having natural environments that are not directly influenced by anthropogenic pollution. These aerosols are assumed to consist of a mixture of 70% water-soluble substances (ammonium and calcium sulphate and also organic components) and 30% dust-like aerosols. The refractive index for the individual components is based on the measurements of Volz⁷. The size distribution is parameterized as the sum of two log-normal size distributions. In contrast, the 'urban aerosols' are considered typical to stagnant polluted air originating from combustion products and industrial sources. This model, therefore, is taken to be a mixture of 80% rural aerosols with 20% carbonaceous aerosols. These soot-like aerosols are assumed to have the same size distribution as both components of the rural model.

To describe aerosol attenuation in finite wavebands, integral turbidity coefficients are often used⁸. For simplicity, we characterize an existing pollution level by its associated surface visibility. Using Koschmieder's law, the visual (meteorological) range $V(\text{km})$ is inversely proportional to the total extinction coefficient at 0.55 μm : $V = 3.91/\sigma$, where $\sigma = \sigma_m + \sigma_a$ (km^{-1}) ('m' and 'a' refer to molecules and aerosols, respectively). In our data base, clear air (that is, $\sigma_a = 0$), corresponds to $V = 300$ km. For increasing pollution levels, the visual range decreases from 300 to 2 km as the aerosol concentration increases from zero to 70,000 particles cm^{-3} for urban aerosols (86,000 particles cm^{-3} for rural aerosols) in the boundary layer up to 2 km height. The particle number densities depend strongly on the relative humidity and the particle size distribution⁵. The values given are for a relative humidity of 90%. In the rest of the atmosphere (up to 70 km height) a constant (light) background pollution level is assumed: the 'tropospheric' aerosol model⁵ up to the tropopause and the 'background stratosphere' model⁵ above the tropopause. This background aerosol is considered to consist largely of sulphate particles formed by photochemical reactions. Aerosol angular scattering is represented by a height-dependent Henyey-Greenstein phase function:

$$P(\theta, z) = \omega_0(z)[1 - g^2(z)][1 - g^2(z) - 2g(z)\cos\theta]^{-3/2} \quad (1)$$

where ω_0 is the aerosol single scattering albedo, and g is the asymmetry parameter. The total phase function is the appropriately weighted average of Rayleigh (molecular) scattering and aerosol scattering as a function of height, z .

Our new calculations use a highly efficient and versatile radiative transfer code⁹ in which we incorporate the detailed atmospheric data base described above¹⁰. The computational method is based on the finite-element discrete-ordinates solution of the radiative transfer equation and has been verified¹¹ against other calculations and actual telephotometer measurements. The computations include the anisotropic scattering characteristics of aerosols, according to equation (1), as well as the detailed vertical distribution profiles for aerosols, molecular scatterers (mainly Rayleigh scattering) and absorbers, as described in the AFGL data. All orders of scattering are taken into account.

We compute the total solar irradiance at ground level for 42 wavelength bins in the photosynthetically active regime from

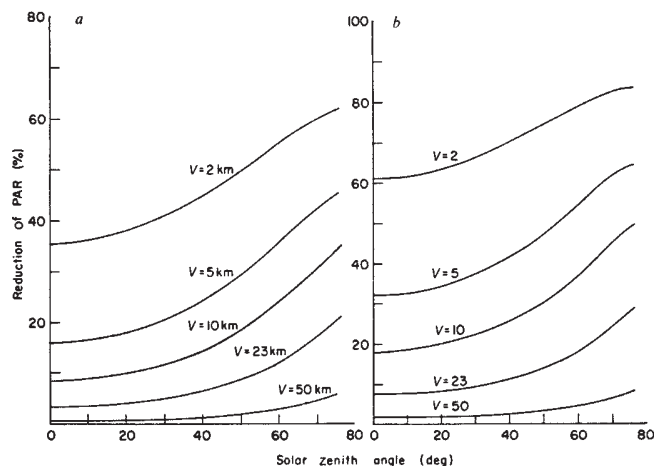


Fig. 1 Per cent reduction of photosynthetically active radiation (PAR) for corn plotted against solar zenith angle. Degree of air pollution is indicated by the surface visual range, V . *a*, Mid-latitude summer atmosphere with rural aerosols in boundary layer. *b*, Mid-latitude summer atmosphere with urban aerosols in boundary layer.

0.35 to 0.72 μm assuming varying aerosol concentrations in the 2 km surface boundary layer. A measure of the biologically effective irradiance is the PAR, defined by

$$\text{PAR} = \int_{\lambda_1}^{\lambda_2} P(\lambda) I(\lambda, Z_0) d\lambda \quad (2)$$

where $I(\lambda, Z_0)$ is the spectral irradiance at ground level for the solar zenith angle Z_0 , and $P(\lambda)$ denotes the photosynthesis action spectrum for a particular plant. Equation (2) broadens the conventional definition of PAR by introducing the weighting function $P(\lambda)$, a convenient way of characterizing the special response of a specific plant species; λ_1 and λ_2 are the wavelengths for which $P(\lambda)$ tends to zero. Photosynthesis action spectra measured for 22 crop plants^{12,13} were found to be fairly similar. Therefore, we report here only our results for corn, which is typical of most agricultural crop plants.

Figure 1 shows the percentage reduction of PAR compared with clear air for five different surface visual ranges that arise when the amount of rural or urban aerosols in the 2 km boundary layer is increased. The increasing values for all curves with increasing solar zenith angle are clearly due to the increased path length for the solar radiation as it passes through the atmosphere at larger angles of incidence. Any fixed solar zenith angle may be thought of as an effective zenith angle, determined as a daily average over the continuously varying direction of the Sun at a given date and latitude¹¹. For example, for the northeastern United States during the summer months, an effective solar zenith angle of 55° is appropriate. In this case we conclude from Fig. 1*a* that if the visual range due to rural aerosols decreases to 10 km, the PAR for corn will be reduced by about 20% compared with clear air conditions. For urban aerosols (Fig. 1*b*) the respective PAR reduction is 33% for the same visibility degradation. In the case of more severe pollution, the calculated PAR reduction increases nonlinearly to very high values, for example, for $V = 2$ km, urban aerosols and $Z_0 = 55^\circ$, we obtain a 75% PAR reduction. As a typical example, a visibility of 20 km (10 km) and more has been observed¹⁴ over the Chicago Metropolitan area for only 40% (70%) of the time between 1970 and 1972.

We note that in the solar radiation measurements of McCartney¹⁵, only small turbidity changes ranging from 0.1 to 0.5 were found. However, when the visual range decreases from 50 to 5 km, the turbidity coefficient (average optical depth, τ) increases from 0.1 to 1.0. The intensity of direct sunlight, which is proportional to $\exp(-\tau)$, therefore decreases by about 60% while the diffuse radiation field due to multiple scattering increases.

The large calculated PAR reductions demonstrated here may appear alarming. Note, however, that any reduction in photosynthetically active radiation cannot be translated directly into an equivalent crop yield reduction because many additional effects in plant physiology and local climatology must be considered also.^{16,17} For example, the so-called C_3 plants reach saturation of their daily photosynthesis at radiation intensities of about one-third full sunlight. Other species, for example the C_4 plants like sugar cane and maize, increase their net photosynthesis continually up to very high irradiance. Also, if constant high air pollution levels are considered, shifts in the main growing season may occur that also affect plant growth and crop yield. Therefore, to estimate quantitatively the influences of our calculated PAR reductions on crop yield, an additional dynamics system study is required for which our model results could be used as input.

We thank Dr Dan Golomb (Center for Energy Policy, MIT) and Dr John Bachmann of the US Environmental Protection Agency for their guidance and technical assistance during the project.

Received 2 April; accepted 16 September 1982.

1. Erickson, P. A. *Environmental Impact Assessment*, 85–204 (Academic, New York, 1979); *Airborne Particles 199* (University Park Press, 1979).
2. Gerstl, S. A. W., Zardecki, A. & Wiser, H. L. *Nature* **294**, 352–354 (1981).
3. Bach, W., Pankrath, J. & Williams, J. (eds) *Interaction of Energy and Climate*, 107–361 (Reidel, Dordrecht, 1980).
4. Jaenicke, R. *J. Aerosol Sci.* **11**, 577–588 (1980).
5. Shettle, E. P. & Fenn, R. W. *Air Force Geophys. Lab. Rep. AFGL-TR-79-0214* (US National Technical Information Service (NTIS accession no. ADA-085-951) (1979); *Proc. AGARD Conf. No 183*, Denmark (NTIS accession no. ADA-028-615) (1975).
6. McClatchey, R. A., Fenn, R. W., Selby, J. E. A., Garing, J. S. & Volz, F. E. *Air Force Cambridge Res. Laboratories, Envir. Res. Pap. 331 AFCRL-70-0527* (NTIS AD-715270) (1970).
7. Volz, F. E. *J. geophys. Res.* **77**, 1017–1031 (1972).
8. Unsworth, M. H. & Monteith, J. L. *Q. J. R. met. Soc.* **98**, 778–787 (1972).
9. Hill, T. R. *Los Alamos nat. Lab. Rep. LA-5990-MS* (1975).
10. Gerstl, S. A. W. *Int. Radiat. Symp.*, Fort Collins, Colorado (1980); *Los Alamos nat. Lab. Preprint LA-UR-80-1403* (1980).
11. Zardecki, A. & Gerstl, S. A. W. *Los Alamos nat. Lab. Rep. LA-9010-MS* (1981).
12. McCree, K. *Agric. Met.* **10**, 443–453 (1972).
13. McCree, K. *Agric. Met.* **9**, 191–216 (1971).
14. Trijonis, J. & Yuan, K. *US Envir. Protection Agency Rep. EPA-600/3-78-075* (1978).
15. McCartney, H. A. *Q. J. R. met. Soc.* **104**, 911–926 (1978).
16. Emerson, R. A. *Rev. Pl. Physiol.* **3**, 1–24 (1958).
17. Lawlor, D. W. in *Stress Physiology in Crop Plants* 303–325 (Wiley-Interscience, New York, 1979).

No significant deviation from random mating of worldwide populations of *Drosophila melanogaster*

N. R. Henderson & D. M. Lambert

Department of Zoology, University of Auckland, Private Bag, Auckland, New Zealand

Multiple choice experiments using 29 populations of *Drosophila melanogaster* from around the world were performed to measure any divergence in the mating system of this species. Courting flies were observed in a total of 691 mating experiments and 9,280 copulations. Only two of the 38 interpopulation crosses exhibited a significant divergence from random mating. However, neither of these represent a trend towards positive assortative mating. This suggests that there is considerable stability in the mate recognition system of *D. melanogaster* in spite of the substantial genetical and morphological differentiation which exists throughout the cosmopolitan distribution of this species.

Components of mating success in *Drosophila* are under genetic control¹ and in laboratory populations the component of darwinian fitness which includes the 'selection' of mates is important in determining reproductive success^{2–4}. Genetically based variation in 'pre-mating isolating mechanisms' occurs among geographically separated populations of a species^{5–7}, either as a result of selection, genetic drift, or both^{5,8,9}. Although much is known about the genetics of reproductive behaviour of *D. melanogaster* in the laboratory, studies have mainly used

Table 1 Results of multiple choice experiments among populations of *Drosophila melanogaster*

Cross	No. of expts	No. of matings				Pairs mated	% Mated	ZI	s.e.
		$\delta 1 \times \varphi 1$	$\delta 2 \times \varphi 1$	$\delta 1 \times \varphi 2$	$\delta 2 \times \varphi 2$				
23 × 28	18	61	48	75	64	248	62	1.04 ± 0.14	
27 × 23	24	76	64	93	73	306	58	0.96 ± 0.10	
28 × 27	23	78	73	85	73	309	61	0.96 ± 0.10	
26 × 24	26	77	116	87	91	371	65	0.83 ± 0.09	*8.9
28 × 26	18	49	54	67	69	239	60	0.97 ± 0.13	
22 × 7	10	27	30	31	39	127	58	1.06 ± 0.17	
27 × 9	18	53	52	61	66	232	59	1.05 ± 0.14	
2 × 15	25	76	82	83	96	337	61	1.04 ± 0.10	
3 × 2	18	71	67	58	53	249	64	0.98 ± 0.12	
3 × 6	18	63	72	59	68	262	66	1.00 ± 0.12	
2 × 28	18	53	56	54	56	218	55	0.99 ± 0.14	
4 × 5	18	66	61	73	71	271	68	1.03 ± 0.12	
22 × 3	18	63	62	56	49	230	58	0.94 ± 0.14	
29 × 14	26	93	85	91	93	362	63	1.06 ± 0.10	
3 × 10	18	64	82	63	68	277	70	0.91 ± 0.13	
13 × 14	18	81	65	66	58	270	68	1.05 ± 0.14	
23 × 19	22	85	87	72	85	329	68	1.07 ± 0.10	
23 × 10	20	70	66	76	81	293	67	1.06 ± 0.10	
19 × 9	22	87	66	78	62	293	60	1.02 ± 0.12	
9 × 15	20	53	51	55	49	208	47	0.96 ± 0.14	
8 × 13	20	56	69	58	70	253	58	1.00 ± 0.12	
11 × 20	9	31	33	27	31	122	62	1.03 ± 0.18	
11 × 12	11	32	36	36	39	143	59	0.98 ± 0.17	
20 × 21	18	60	54	52	50	216	55	1.03 ± 0.14	
25 × 17	20	63	59	70	71	263	60	1.04 ± 0.12	
18 × 27	10	45	44	38	40	167	75	1.04 ± 0.15	
22 × 15	10	32	38	34	32	136	62	0.89 ± 0.16	
27 × 7	22	78	75	75	73	301	62	1.01 ± 0.10	
4 × 14	15	37	37	53	61	188	57	1.07 ± 0.16	*9.1
25 × 22	24	70	66	93	81	310	59	0.96 ± 0.11	
15 × 16	17	52	43	55	53	203	54	1.07 ± 0.15	
5 × 1	17	57	60	57	60	234	63	1.00 ± 0.14	
9 × 10	22	77	75	73	74	299	62	1.02 ± 0.10	
21 × 23	23	75	67	91	86	319	63	1.02 ± 0.10	
13 × 6	22	62	52	78	65	257	53	1.00 ± 0.12	
19 × 14	11	37	34	39	33	143	59	0.96 ± 0.16	
12 × 21	11	39	39	35	36	149	62	1.01 ± 0.16	
8 × 2	11	39	37	34	36	146	60	1.05 ± 0.17	

The experiments were performed at 25 °C by direct observation, a technique originally devised by Elans and Wattiaux¹⁰. For each experiment 11 virgin males and females of one strain, plus 11 virgin males and females of a second strain, a total of 22 pairs, were used for every set of two strains. Four types of mating could thus be observed:

$$\delta\delta 1 \times \varphi\varphi 1, \delta\delta 2 \times \varphi\varphi 1, \delta\delta 1 \times \varphi\varphi 2, \delta\delta 2 \times \varphi\varphi 2.$$

Flies were kept separately for 5–7 days and then introduced into a mating chamber and observed at 6-min intervals for 1 hour. This duration allows for about 60% of the females to mate. Records were made of the number of within and between strain matings. In order to distinguish the flies of the different strains through the transparent cover of the observation chamber, a small dot was painted on the dorsal surface of the thorax with non-toxic ink (Rotring). This marker was rotated from strain to strain, so that no one strain in any combination of two strains tested together, always had the thorax painted. Painting was performed on lightly etherized flies at the same time as they were 'sexed' after emergence. Marking was found to have no effect on the strains studied here. All experiments were conducted between 0900 and 1200 hours. All experiments for any one cross were run on the same day. For analysis of the data, comparison with panmixia the χ^2 test was used. In addition isolation was measured by Levenes joint index^{11,12}. The number of males and females of the two strains being equal, let the observed number of matings of the four possible types be given by:

	Strain 1 (♂♂)	Strain 1 (♀♀)	Strain 2 (♀♀)
Strain 1 (♂♂)	X11	X12	
Strain 2 (♂♂)	X21	X22	

Levene defines the Joint isolation index as:

$$ZI = \sqrt{\frac{X11 \cdot X22}{X12 \cdot X21}}$$

ZI = 1 Means random mating. ZI > 1 Means a tendency to homogamic mating. ZI < 1 Means a tendency to heterogamic mating.

The exact distribution of these statistics is unknown but practical tests of significance can be based on the following estimates of variance:

$$\sigma^2(ZI) = \frac{(ZI)}{(4)} H \text{ where } H = \frac{1}{X11} + \frac{1}{X12} + \frac{1}{X21} + \frac{1}{X22}$$

* $P < 0.05$, 3d.f.

mutant strains or well established wild-type cultures. We know surprisingly little about the genetic dynamics of the mating systems of natural populations. In order to test for any geographic differences in the male-female communication system, multiple choice experiments¹⁰ were conducted among 29 populations derived from diverse localities around the world (Fig. 1). Wherever possible each population was derived from a large number of recently collected individuals, and populations were kept for only a short time before use. Thirty-eight crosses between localities were undertaken and results were tested for deviations from random mating using a chi-squared test (Table 1). This does not, however, indicate the degree and nature of the mating preference should a significant result appear and for this purpose the degree and nature of any mating divergence was measured using Levenes joint isolation index¹¹.

The data show no significant deviation from random mating.

Only two crosses resulted in a significant deviation and neither of these are the result of a tendency to mate within populations. The cross between the two Russian populations 26 and 24, for example, is significant because males of population 24 have a higher propensity to mate. The remaining cross which yielded a significant result, cross 4 × 14, was the result of a higher mating propensity on the part of females of population 14.

Previous studies of natural populations of *D. melanogaster* involved crosses between French, Japanese and Afrotropical strains^{12,13}, also showed a lack of divergence. Our techniques were very similar so we propose that this is a general characteristic of populations of *D. melanogaster*. There is a similar lack of divergence in mating behaviour of populations of *D. pseudoobscura*, from western United States and Mexico¹⁴. The latter authors point out, however, that the opposite result is common in interpopulation crosses involving *Drosophila*

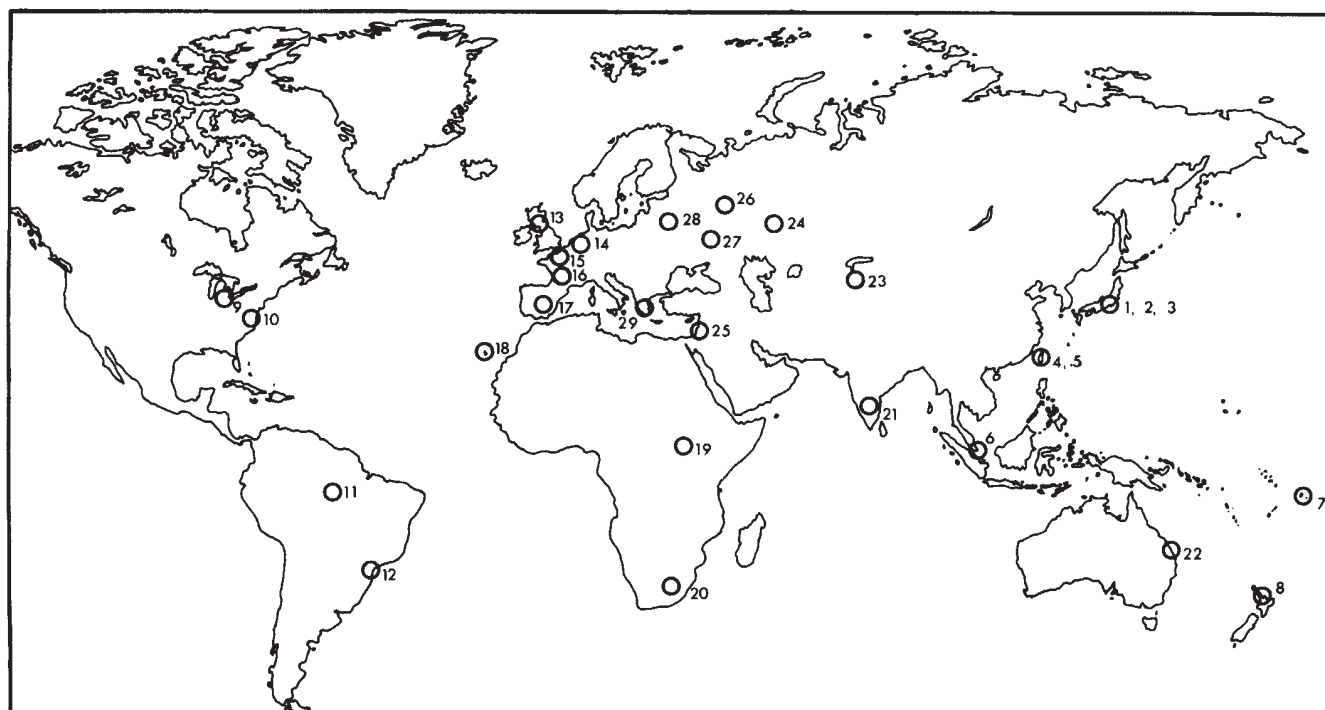


Fig. 1 Collection localities of strains of *D. melanogaster* used in the crossing experiments. The exact location, numbers of individuals from which the strains were derived, and date of collection are listed in parentheses. (1) Mishima, Japan; >10♀; August 1981; (2) Katsunuma, Japan; 50♀; October 1981; (3) Otsuki City, Japan; 8♀; September 1981; (4) Wulai, Taipei County, Taiwan; 1♀; November 1981; (5) Tong Pu, Nan-Tou County, Taiwan; 1♀; November 1981; (6) Singapore; large number of adults; 1979; (7) Suva, Fiji; 4♀; August 1981; (8) Auckland New Zealand; 20♀; February 1982; (9) Bowling Green, Ohio; large number of adults; October 1981; (10) Beltsville, Maryland; large number of adults; September 1981; (11) Jose Bonifacio (Sao Paulo) Brasil; 10♀ 8♂; September 1981; (12) Manaus (Amazonas) Brasil; 2♀ 3♂; November 1981; (13) Edinburgh; 20 adults; July 1981; (14) Gröningen, Holland; >30 adults; July 1981; (15) Ménétiéol (Sancienne) France; 28♀ 112♂; September 1981; (16) Iseré (Lyon) France; 200 adults; September 1981; (17) 70km from Madrid, Spain; 400 adults; November 1981; (18) Tenerife, Canary Islands; 25♀; September 1981; (19) Kampala, Uganda; 10♀; October 1981; (20) Pretoria, South Africa; 15♀; February 1982; (21) Nagarhoe, 15km Mysore, India; 11♀; February 1982; (22) Frazer Isle, Queensland, Australia; 50♀; September 1981; (23) Alma-Ata, Soviet Union; small number of adults; Mid-1981; (24) Birk (50 miles from UFA); small number of adults; Mid-1981; (25) Qiryat Anavim (Jerusalem), Israel; 50–100 adults; September 1981; (26) Gorky, Soviet Union; small number of adults; Mid-1981; (27) Voronez, Soviet Union; small number of adults; Mid-1981; (28) Vitebsk, Soviet Union; small number of adults; Mid-1981; (29) Gavros-Achaia, Greece; >25♀; September 1981.

species. They list 21 previous studies, 19 of which cite data showing significant variation in mate recognition. However, they have a number of reservations regarding this result and it remains whether this is a general characteristic of natural populations or a result of experimental technique.

Our results can be interpreted in two possible ways. First, recent dispersion of *D. melanogaster* may have allowed little time for the development of genetic differences between populations. Alternatively, populations of *D. melanogaster* have indeed differentiated genetically in elements of their biology, but stabilising selection acting on the reproductive communication system has prevented it from diverging.

If there is an overall lack of divergence then our results are not surprising. However, recent research on the biology of natural populations of *D. melanogaster* has revealed considerable genetic differences, including biometry^{16–20}, electromorphs^{21,22}, chromosomal inversion frequencies^{23,24}, morphological mutant phenotypes²⁵, ecology^{26,27}, physiology¹⁹. These differences may be the result of selection in different local conditions. Recent investigations of hybrid dysgenesis^{28–30} in *D. melanogaster* indicate that potential for 'postmating reproductive isolation' exists among geographically diverse populations. Why then have the genes controlling mating behaviour apparently resisted such change? Deviation in mating behaviour of laboratory populations of *D. melanogaster* has been recorded by Kiliass *et al.*³¹. However, this may be due to non-genetic changes in mating rhythms which if corrected for, may result in no significant deviation for example ref. 32.

The second interpretation of our data is that stabilizing selection exerted through the interactions of the organisms themselves, acts on the reproductive communication system to maintain its stability. The function³³ of this male–female communication system is to facilitate the achievement of efficient

syngamy in the conditions prevailing in the habitat of the organisms³⁴. It has been argued (see ref. 33 for a recent view), that individuals with a deviant mate recognition system, would obtain less matings than those members of the population performing the correct sequences^{34–36} that are necessary for successful reproduction. These deviant individuals would thus be less fit. It might be expected then that the mate recognition system should show a degree of geographic stability³⁷ and the data presented here support this prediction.

We are grateful to the following workers who sent us wild caught material: J. D. Agnew, S. M. Akai, S. Alahiotis, H. M. Bicudo, W. W. Cantello, A. W. Ewing, R. Falk, A. Fleuriot, M. D. Golubovsky, A. M. Gonzales, M. Gromko, S. N. Hegde, W. Kerver, E. R. Kiberu, O. Kitagawa, F.-J. Linn, C. Monchamp-Moreau, J. Oakeshott, M. D. Ochando, C. Petit, V. P. E. Phang, C. Postlethwaite, D. Van-Delden, S. Verma, T. K. Watanabe. We would also like to thank E. Kay, A. Oliver, M. Soochoon for technical assistance and B. Philip, M. C. McLea, A. A. Harper, J. Craig and E. C. Young for helpful comments on the manuscript. The work was aided by funds from the Auckland University Grants Committee, grant no.141Z117 (to D.M.L.), and the Department of Zoology, University of Auckland.

Received 8 September; accepted 7 October 1982.

1. Speiss, E. B. in *Essays in Honour of Theodosius Dobzhansky* (eds Hecht, M. K. & Steere, W. C.) 315–379 (North Holland, Amsterdam, 1970).
2. Bongaard, J. & Christianson, F. B. *Genetics* **71**, 439–460 (1972).
3. Parsons, P. A. *Behav. Genet.* **4**, 395–404 (1974).
4. Merrell, D. J. *Evolution* **7**, 287–296 (1953).
5. Mayr, E. *Animal Species and Evolution* (Harvard University Press, Cambridge, 1963).
6. Dobzhansky, T. in *Molecular Evolution* (ed. Ayala, F. J.) 95–105 (Sinauer, Sunderland) Massachusetts (1976).
7. White, M. J. D. *Modes of Speciation* (Freeman, San Francisco 1978).
8. Parsons, P. A. *Adv. Genet.* **19**, 1–32 (1977).
9. Parsons, P. A. *Am. Nat.* **111**, 613–621 (1977).
10. Elans, A. A. & Wattiaux, J. M. *Drosoph. Inf. Serv.* **39**, 118–119 (1964).

11. Ehrman, L. & Petit, C. *Evolution* **22**, 649–658 (1968).
12. Petit, C., Kitagawa O. & Takamura, T. *Jap. J. Genet.* **51**, 99–108 (1976).
13. Cohet, Y. & David, J. R. *Genetica* **54**, 161–165 (1980).
14. Anderson, W. W. & Ehrman, L. *Am. Midl. Nat.* **81**, 47–53 (1969).
15. Dobzhansky, T. in *The Genetics of Colonising Species* (eds Baker, H. G. & Stebbins, G. L.) (Academic, New York 1965).
16. Teissier, G. *Annls Genet.* **1**, 2–10 (1958).
17. Petit, C. *Bull. Soc. Zool. Fr.* **94**, 641–648 (1969).
18. David, J. R. & Bocquet, C. *Experientia* **31**, 164–166 (1975).
19. David, J. R. & Bocquet, C. *Nature* **257**, 588–590 (1975).
20. Takanashi, E. & Kitagawa, O. *Drosoph. Inf. Serv.* **52**, 28–29 (1977).
21. Girard, P., Palabost, P. L. & Petit, C. *Biochem. Genet.* **15**, 589–599 (1977).
22. Oakeshott, J. G., Chambers, S. K., Gibson, J. B. & Willcocks, D. A. *Heredity* **47**, 385 (1981).
23. Stalker, H. D. *Genetics* **82**, 323–347 (1976).
24. Stalker, H. D. *Genetics* **95**, 211–223 (1980).
25. Ochando, D. *Experientia* **34**, 833–834 (1978).
26. Parsons, P. A. *Evolution* **33**, 131–136 (1979).
27. Marks, R. W., Brittnacher, J. G., McDonald, J. F., Prout, T. & Ayala, F. J. *Oecologia (Berl.)* **47**, 141–144 (1980).
28. Kidwell, M. G. *Genetics* **98**, 275–290 (1981).
29. Kidwell, M. G. in *Genetics and Biology of Drosophila* Vol. 3 (eds Ashburner, M., Carson, H. L. & Thompson, J. N. Jr) (Academic, New York, 1982).
30. Kidwell, M. G., Kidwell, J. F. & Sved, J. A. *Genetics* **86**, 813–833 (1977).
31. Kilias, G., Alahiotis, S. N. & Pelecanos, M. *Evolution* **34**, 730–737 (1980).
32. Harper, A. A. *Drosoph. Inf. Serv.* (in the press).
33. Lambert, D. M., Kingett, P. D. & Slooten, E. *Evol. Theory* (in the press).
34. Bastock, M. & Manning, A. *Behaviour* **8**, 85–111 (1955).
35. Paterson, H. E. *Evolution* **34**, 330–331 (1980).
36. Paterson, H. E. *S. Afr. J. Sci.* **77**, 113–119 (1981).
37. Lambert, D. M. & Paterson, H. E. *Evol. Theory* **5**, 291–300 (1982).

Birth sex ratios and social rank in rhesus monkey mothers

M. J. A. Simpson & A. E. Simpson

MRC Unit for the Development and Integration of Behaviour,
Madingley, Cambridge CB3 8AA, UK

The ability to control the sex of her offspring could be of survival value to a mother^{1,2}, and prompts questions about mechanisms of sex determination. Trivers and Willard have suggested that mothers potentially able to invest much in offspring should bear those kinds of infant that repay high levels of maternal investment most effectively². One hypothesis concerning matrilineal primate species states that a dominant mother would leave more descendants through a daughter rather than a son if that daughter inherited her high rank^{3–5}. On the other hand, a low ranking mother of such a species would leave more descendants through a son, for he is likely to emigrate at puberty and not necessarily inherit his mother's low rank⁶. Unfortunately, long-term data on primate birth sex ratios are in short supply. Here we present data collected over 20 years on a captive colony of rhesus monkeys which support the above hypothesis. High ranking rhesus monkey mothers in our captive colony⁷ were more likely to give birth to daughters than sons, while the remaining mothers were more likely to bear sons than daughters. Finally, we suggest a mechanism for adjusting birth sex ratios in macaques.

The conceptions and pregnancies referred to in this study occurred in six small social groups of rhesus monkeys (*Macaca mulatta*). Each group was housed in its own indoor room with adjoining outdoor run, at the Subdepartment of Animal Behaviour, near Cambridge⁷. There were 7–12 monkeys in each group—one adult male, two to five adult females (at least 4 yr old), and immature offspring born into that group. The groups shared the same building so that every monkey could see and hear every other one. As far as possible, families of mothers and their female offspring were kept in their original groups. We removed young males before their fifth years, and changed incumbent adult males often enough to prevent them from impregnating their daughters. In this paper our data exclude pregnancies resulting from matings arranged with males held apart from the mother's home group. Also excluded are pregnancies spent partly away from the home group.

Rhesus monkey mothers having high social ranks gave birth to twice as many daughters as sons (Table 1 (1960–82), $P = 0.0026$, binomial test⁸), while the remaining mothers bore more sons than daughters ($P = 0.024$, binomial test). (Among the non-high ranking mothers, middle and low ranking ones did

Table 1 Number of infants of each sex born to mothers of different ranks

Births between:	1960–71		1972–81		1960–81	
	Daughters	Sons	Daughters	Sons	Daughters	Sons
High rank mothers	18	6	20	9	38	15
Other mothers	7	15	25	39	32	54

A total of 53 mothers and 139 births are included in the sample. All conceptions and pregnancies occurred in the social group. Eight infants dying between their first and fourteenth days, and all available (16) sexed stillbirths are included. Social ranks of mothers living in small captive groups⁷ were assessed before we knew of their possible connection with birth sex ratios. In practice, a high ranking mother is easily recognized as dominating the other females in her group. To represent the mother's rank at the likely time of conception, we used her rank when the resulting infant was 52 weeks old, for in only 20 cases did we have information about the mother-to-be's rank at conception. In none of these cases was there a change between the two rank positions. Information about the ranks of mothers with younger infants was not used because their greater dependence could have influenced their mothers' ranks. From 1960 to 1971, assessments of rank were made by earlier observers²². Thereafter, attacking, chasing, displacing, avoiding, fear grinning, and presenting (non-sexual, non-grooming context) between a pair of monkeys were regarded as evidence for dominance. In any dyad, one monkey was deemed dominant to the other if their actions occurred three or more times as often in one direction as in the other. A mother's rank reflects the direction but not necessarily the rate of aggressive behaviour between herself and others. A high ranking female would not be aggressive so long as the others avoided her, but the others could be fighting among themselves at a high rate.

not differ in their birth sex ratios.) The effect of maternal dominance on birth sex ratio was consistent from the 11-yr period preceding 1972 to a subsequent 10-yr period. Most mothers contributed more than one infant to Table 1, but when each mother was classified according to whether she bore more daughters than sons or vice versa, the contrast in birth ratios between high ranking and other mothers was still apparent (Table 2, $P = 0.03$, 0.03 and 0.001 for the three time periods, Fisher exact probability test). The effect also shows in first births to the mothers (Table 3, $P =$ not significant, 0.02 and 0.10 for the three time periods, Fisher exact test), suggesting that it could be independent of a mother's previous reproductive history.

Altmann³ showed the same effect in free-living yellow baboons (*Papio cynocephalus*), her ratios being 18:10 daughters to sons for high ranking mothers, and 8:15 for the remainder. Silk's⁴ low ranking (captive) bonnet macaques (*Macaca radiata*) gave birth to 20:44 females to males, although females from the top half of the hierarchy did not have more daughters than sons.

Until we know more about the mechanisms controlling birth sex ratios and the ways in which social factors operate¹, hypotheses about the adaptive significance of birth sex ratios

Table 2 For mothers holding only one rank across all their infants, the number of mothers of each rank who produced more daughters than sons, or sons than daughters

	1960–71		1972–81		1960–81*	
	More daughters	More sons	More daughters	More sons	More daughters	More sons
High rank mothers	7	2	5	0	9	0
Other mothers	3	10	5	10	8	17

Some mothers contributed infants to both time periods, and 13 mothers who changed rank were excluded.

* Six mothers (three high, three other) had equal numbers of daughters and sons.

must remain tentative, for the mechanism acting in our colony may be triggered by as yet unspecified special conditions in it. Meanwhile, we believe that our socially living mothers bias the birth sex ratios of their offspring according to some aspect of their dominance rank. Such a mechanism would have survival value for high ranking rhesus monkey mothers if daughters best repaid their mothers' high levels of maternal investment in them. This is a general form of the Trivers and Willard hypothesis² that mothers potentially able to invest much in offspring should bear and rear the kinds of infants repaying the highest levels of maternal investment most effectively. Such infants could be daughters in some socially living macaque and baboon species^{3,4}, although they were believed to be sons in the cases considered by Trivers and Willard.

In support of our view, rhesus monkey daughters remain with their mothers and inherit their mothers' social ranks⁵, so that a high ranking daughter could contribute more to her mother's fitness by aiding in interfamily fights⁹ and by breeding more successfully¹⁰ than a son, who is likely to emigrate at puberty¹¹⁻¹³, and not necessarily achieve high social rank in the group he subsequently joins⁶.

However, daughters bring costs as well as benefits to their mothers, for example by competing with each other within their families, and 1- and 2-yr-old females in our colony are especially likely to behave with jealousy and aggression to their mothers and subsequently born siblings¹⁴. Emigrating sons of high ranking mothers could take with them certain advantages gained

hypothesis could not explain our bias towards female births in high ranking mothers.

We suggest that rhesus monkey mothers, like white-tailed deer¹⁸, could adjust the sex ratios of their infants if they controlled the timing of mating relative to ovulation. At present we know of no supporting evidence from non-human primates, although in human beings weak biases towards males have occurred when zygotes are formed early or late in the fertile period rather than in the middle¹⁹, and biases towards females occurred when gonadotropin levels were deliberately made high²⁰. The question of how the timing of female sexual behaviour close to the time of ovulation depends on the mother's social situation²¹ remains open.

This research was supported by the MRC. We thank R. A. Hinde for his comments on the manuscript.

Received 6 July; accepted 6 October 1982.

1. Clutton-Brock, T. H. in *Current Problems in Sociobiology* (Cambridge University Press, in the press).
2. Trivers, R. L. & Willard, D. E. *Science* **179**, 90-92 (1973).
3. Altmann, J. *Baboon Mothers and Infants* (Harvard University Press, Cambridge, Massachusetts, 1980).
4. Silk, J. B., Clark-Wheatley, C. B., Rodman, P. S. & Samuels, A. *Anim. Behav.* **29**, 1106-1120 (1981).
5. Sade, D. S. in *Social Communication Among Primates* (ed. Altmann, S. A.) 99-114 (University of Chicago Press, 1967).
6. Koford, C. B. *Science* **141**, 336-337 (1963).
7. Anderson, D. M. & Simpson, M. J. A. *Lab. Anim.* **13**, 275-281 (1979).
8. Siegel, S. *Nonparametric Statistics* (McGraw-Hill, New York, 1956).
9. Kaplan, J. R. *Am. J. phys. Anthropol.* **49**, 241-249 (1978).
10. Chapais, B. & Schulman, S. *J. theor. Biol.* **82**, 47-89 (1980).
11. Lindburg, D. G. in *Primate Behavior 2* (ed. Rosenblum, L. A.) 1-106 (Academic, New York, 1971).
12. Packer, C. *Anim. Behav.* **27**, 1-36 (1979).
13. Drickamer, L. C. & Vessey, S. H. *Primates* **14**, 359-368 (1973).
14. Spencer-Booth, Y. *Anim. Behav.* **16**, 541-557 (1968).
15. Drickamer, L. C. *Folia Primatol.* **21**, 61-80 (1974).
16. Simpson, M. J. A., Simpson, A. E., Hooley, J. & Zunz, M. *Nature* **290**, 49-51 (1981).
17. Sackett, G. P., Holm, R. A., Davis, A. E. & Fahrenbruch, C. E. *Symp. 5th Cong. int. Primat. Soc.* (eds Kondo, S., Kawai, M., Ehara, E. & Kawamura, S.) 189-205 (Japan Science, Tokyo, 1975).
18. Verme, L. J. & Ozoga, J. J. *J. Wildl. Mgmt* **45**, 710-715 (1981).
19. James, W. H. *Lancet* **i**, 1124-1128 (1980).
20. James, W. H. *Br. med. J.* **281**, 711-712 (1980).
21. Hausfater, G. *Contrib. Primatol.* **7**, 1-150 (1975).
22. Richards, S. M. *Anim. Behav.* **22**, 914-930 (1974).

Table 3 Number of infants of each sex born to primiparous mothers of different ranks

	1960-71		1972-81		1960-81	
	Daughters	Sons	Daughters	Sons	Daughters	Sons
High rank mothers	1	2	4	0	5	2
Other mothers	1	1	3	11	4	12

In the second time period, while 11 'other' mothers had sons as their first infant, only 10 went on to have more sons than daughters (see Table 2).

from their mothers' statuses even if they did not retain their mothers' ranks.

Functional arguments can be applied to low ranking mothers also. A daughter of a low ranking mother may share with her mother the disadvantages of low rank, including poorer breeding performance¹⁵, while an emigrating son, whose rank is not thereafter tied to his mother's, may gain sufficient rank in his new social group to contribute considerably to his mother's fitness. In addition, young daughters may be more costly to rear than sons, being more involved in family strife, and bringing more aggression to their mothers from other adults⁴. Mothers were also slower to breed again after bearing daughters rather than sons¹⁶.

However, these functional arguments depend partly on results from other social groups and species. To assess the balance of costs and benefits for high and low ranking mothers rearing sons and daughters, we need to know both the costs and the benefits to individuals through the reproductive lives of themselves and their offspring¹⁰.

The mechanisms whereby these sex ratios are controlled is unknown¹. Differential effects of stress on the proportions of XX- and XY-bearing sperm reaching the ovum³, or on the post-conception survival of male and female fetuses¹⁷ would explain our bias away from daughters in the births to low ranking mothers, if XX-sperm and/or female fetuses were more vulnerable, and low rank were associated with greater stress. Potentially stressful situations, however, did not increase the chance that our mothers would bear sons: 10 mothers scoring above the median on both aggression received from other adults and submissive behaviour given to others (see Table 1) did not give birth to more sons than daughters. Moreover, a stress

Use of the UCHT1 monoclonal antibody to explore mouse mutants and development

K. W. T. Caddy, D. L. Patterson & T. J. Biscoe

Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

Garson and Coakham demonstrated a monoclonal antibody (UCHT1) which specifically labels Purkinje cells in the cerebellar cortex¹. To investigate mouse mutants with cerebellar lesions, we obtained some of the purified antibody from Beverley², who derived it as an antibody to a human T-cell lymphocyte surface antigen. We have used the indirect immunoperoxidase method to label with the antibody Purkinje cells in mutant and normal mice, frog, toad, goldfish, dogfish and 3-acetylpyridine-treated rat. We report here that in the mutant mice tested, only the Purkinje cells in *staggerer* did not stain. The Purkinje cells of frog, toad and goldfish were also unstained but those in the dogfish were labelled. The Purkinje cells in rats which had been injected with 3-acetylpyridine to destroy the climbing fibres, were found to stain with the antibody. These results suggest phylogenetic as well as ontogenetic differences in antigenicity of Purkinje cells.

The cerebella in the mutant mice used all have different, well defined lesions: *stumbler* Purkinje cells have stunted dendritic trees³; *Lurcher* loses all its Purkinje cells in the first 60 days after birth⁴; *weaver* loses all of its granule cells and has abnormal Purkinje cell dendritic trees⁵; *reeler* has a disorganized

cerebellar cortex with Purkinje cells in abnormal positions⁶, and *staggerer* has reduced numbers of Purkinje cells⁷.

Mutant and littermate control animals were all aged between 15 and 20 days postnatal when used for these initial experiments because *Lurcher* begins to lose significant numbers of Purkinje cells after 15 days. The *stumbler*, *Lurcher*, *weaver* and *reeler* mutants all showed positive staining of Purkinje cells with UCHT1. In all cases the whole of the Purkinje cell was labelled; the cell body (excluding the nucleus), the dendritic tree, and the axon, which could be traced in some cells all the way to the deep cerebellar nuclei. With this labelling, the abnormal Purkinje cell dendritic trees were seen as described previously in Golgi preparations^{3,4,8,9}.

In the *staggerer* mutant, however, the Purkinje cells were completely unstained. Figure 1*a, b* shows sections of a control and *staggerer* cerebellum stained with UCHT1 using the indirect immunoperoxidase technique. These sections were not counterstained and were photographed using Nomarski optics to allow Purkinje cells unlabelled by antibody to be seen in the *staggerer* section (Fig. 1*b*).

There was a difference in staining between the control and *stumbler* Purkinje cells. In control animals the staining was

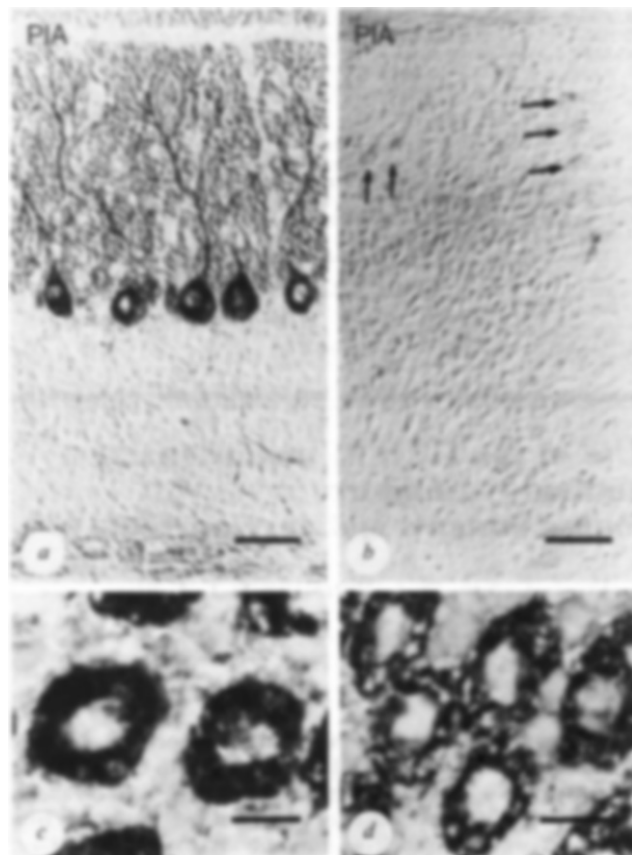


Fig. 1 *a, b*, Sagittal sections of cerebellum of 20-day-old normal and *staggerer* mice, respectively, stained by UCHT1 using the indirect immunoperoxidase technique. Cryostat sections (6 μm) were fixed for 5 min in acetone, incubated for 30 min with a purified IgG fraction of the UCHT1 ascites diluted 1:1,000, washed for 5 min in Tris-buffered saline, and then incubated for 30 min at 20°C with HRP-conjugated rabbit anti-mouse immunoglobulin (DAKO; 1:100). The peroxidase was visualized using 0.6 mg ml⁻¹ diaminobenzidine. Sections were photographed with a Leitz orthoplan microscope and Vario-orthomat camera using Nomarski optics. Arrows in *b* indicate unstained Purkinje cells. Scale, 50 μm . *c, d*, Sagittal sections of Purkinje cells from 19-day-old normal and *stumbler* mice, respectively, stained with UCHT1 and counterstained with haematoxylin. Note that the HRP appears more clumped with numerous small clear areas in the *stumbler* Purkinje cells because of the increased numbers of mitochondrial profiles in these cells. Scale bar, 10 μm .

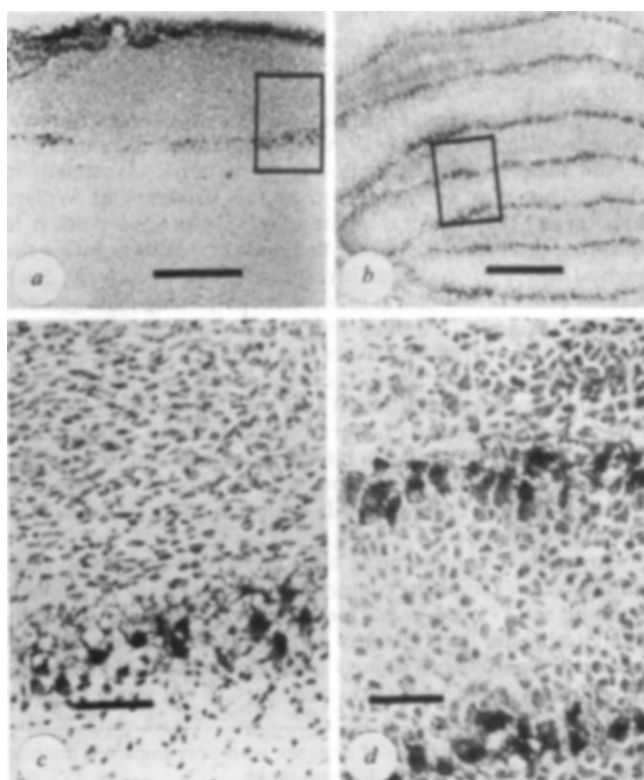


Fig. 2 *a, b*, Horizontal 6- μm sections of 2- and 3-day-old normal mice cerebella stained with UCHT1 and counterstained with haematoxylin. Note the patchy staining of Purkinje cells in *a*. Low power bright field; scale bar, 500 μm . *c, d*, Areas within rectangles in *a* and *b* shown at higher magnification. Scale bar, 50 μm .

uniform whereas in *stumbler* the horseradish peroxidase (HRP) gave a clumped appearance with numerous small clear areas (Fig. 1*c, d*). These abundant clear areas are almost certainly due to the large numbers of mitochondria present in the soma of *stumbler* Purkinje cells³.

We also studied the labelling pattern of normal mice aged 0 (birth), and at days 1, 2, 3, 5 and 8 postnatal. At time 0 and on day 1 the Purkinje cells were found not to stain with UCHT1 but at day 2, patches of stained Purkinje cells were clearly visible (Fig. 2*a, c*). At day 3 (Fig. 2*b, d*) and for all subsequent ages, the Purkinje cells stained as in the day 20 normal control, except that the Purkinje cells at less than 8 days old were noticeably less mature.

We know that between 2 and 3 days postnatal, the age at which the particular antigen appears, the mouse climbing fibres first start to form synapses on the Purkinje cells via somatic spines (K.W.T.C., unpublished observations). Two studies^{10,11} indicate that although the synapses of the climbing fibres are present in the *staggerer*, they are not functional except after injection of harmaline; this suggests that the presence of climbing fibres may be correlated with the presence of the antigen. To examine this hypothesis, we treated rats with 3-acetylpyridine, harmaline and nicotinamide¹² to destroy selectively the inferior olivary nuclei, which are believed to be the sole source of climbing fibres¹³. Our experiments showed no change in the Purkinje cell labelling in any part of the cerebellar cortex. We conclude that the continued presence of climbing fibres is not a requirement for the presence of the antigen in Purkinje cells.

The possibility remains that the appearance of the antigen correlates with the appearance of the climbing fibre innervation. This would explain our findings with the young animals but not the results with the *staggerer* mutant. In *staggerer* it has been shown that the development of the Purkinje cells is retarded¹⁴ to the extent that climbing fibre synapses are still present on somatic spines of animals 23 days after birth. It has also been

shown that *staggerer* Purkinje cells have no parallel fibre-dendritic spine synapses¹⁵. This failure to form branchlet spines in *staggerer* is almost certainly due to some inherent defect in the Purkinje cells since these cells have been shown to be the site of action of the mutant gene¹⁶. Whether or not the absence of the antigen in *staggerer* is causally related to the *staggerer* disorder is an open question.

To determine any further possible correlation between the presence of the antigen and the structure of the cerebellum, we have extended the survey of the animal kingdom begun by Garson *et al.*¹. The species tested were frog (*Rana temporaria*), toad (*Rana esculenta*), goldfish (*Carassius auratus*), dogfish (*Scyliorhinus canicula*) and pigeon. We used the indirect immunoperoxidase method to label Purkinje cells in at least two of each of these species. Although this study is at best preliminary, it is of interest to note that only in the dogfish and pigeon did the Purkinje cells stain. This finding is interesting because Llinas and Hillman¹⁷ have shown that the cerebellum of the elasmobranchs is more complex than that in frog and

other amphibia. By this they mean that there are fewer cell types in the frog (basket cells missing) and the Purkinje cells are less complex. In fact, the Purkinje cells in the frog have far fewer spiny branchlets on their dendrites than elasmobranchs and probably present a more primitive form of cerebellar cortex¹⁷. The similarity in the gross morphology between *staggerer* and frog Purkinje cells may be of importance in this context.

It seems that the antigenic determinant for UCHT1 is absent from some species, from at least one mutant mouse and also from mice less than 2 days old. At this stage it is impossible to correlate the antigen with any structure in Purkinje cells but we hope that further electron microscope and biochemical studies may elucidate its location.

We thank Drs Garson and Coakham; Dr P. Beverley of the Imperial Cancer Research Fund, University College Hospital for the gift of the UCHT1; N. Bowery, A. Warner, G. Falk and M. Whitear for supplying the *staggerer*, *reeler* and *weaver* mutant mice, toad, dogfish and goldfish. This research was supported by Action Research.

Received 21 July; accepted 8 October 1982.

- Garson, J. A., Beverley, P. C. L., Coakham, H. B. & Harper, E. I. *Nature* **298**, 375–377 (1982).
- Beverley, P. C. L. & Callard, R. E. *Eur. J. Immun.* **11**, 329–334 (1981).
- Caddy, K. W. T. & Sidman, R. L. *Dev. Brain Res.* **1**, 221–236 (1981).
- Caddy, K. W. T. & Biscoe, T. J. *Phil. Trans. R. Soc.* **B287**, 167–201 (1979).
- Sotelo, C. *Brain Res.* **94**, 19–44 (1975).
- Hamburgh, M. *Dev. Biol.* **8**, 165–185 (1963).
- Herrup, K. & Mullen, R. J. *Brain Res.* **172**, 1–12 (1979).
- Bradley, P. & Berry, M. *Brain Res.* **142**, 135–141 (1978).
- Mariani, J., Crepel, F., Mikoshiba, K., Changeaux, J. P. & Sotelo, C. *Phil. Trans. R. Soc.* **B281**, 1–28 (1977).
- Crepel, F. & Mariani, J. *Brain Res.* **98**, 135–147 (1975).
- Crepel, F., Delhaye-Bouchard, N., Guastavino, J. M. & Sampaio, I. *Nature* **283**, 483–484 (1980).
- Llinas, R., Walton, K., Hillman, D. E. & Sotelo, C. *Science* **190**, 1230–1231 (1975).
- Armstrong, D. M. *Physiol. Rev.* **54**, 358–417 (1974).
- Hirano, A. & Dembitzer, H. M. *J. Neuropath. exp. Neurol.* **34**, 1–11 (1975).
- Landis, D. M. D. & Sidman, R. L. *J. comp. Neurol.* **179**, 831–863 (1978).
- Herrup, K. & Mullen, R. J. *Brain Res.* **178**, 443–457 (1979).
- Llinas, R. & Hillman, D. E. *Neurobiology of Cerebellar Evolution and Development* (ed. Llinas, R.) 43–73 (American Medical Association, 1969).

Meiotic crossing-over between the X and Y chromosomes of male mice carrying the sex-reversing (*Sxr*) factor

E. P. Evans* & M. D. Burtenshaw

Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE, UK

B. M. Cattanaach

M.R.C. Radiobiology Unit, Harwell, Didcot OX11 0RD, UK

In the mouse, as in many other species, the X and Y chromosomes are seen to pair end-to-end at diakinesis¹, which has led to speculation as to whether chiasmata and crossing-over occur earlier in meiosis¹. Arguments in favour of this have recently been presented^{2–4}, but because of the absence of appropriate genetic and cytological markers little clear-cut evidence has been obtained. Studies with 'sex-reversed' (*Sxr*) mice⁵ carried out earlier this year⁶ have suggested that the transfer of chromatin from the Y to the X chromosome takes place during meiosis, but in this instance the interpretation was confused by the claim that X/Y *Sxr* mice possess two Y chromosomes, contrary to earlier reports⁷. We present evidence here that: (1) X/Y *Sxr* mice have a single abnormal Y chromosome possessing an additional dark-staining distal body; (2) during meiosis, this body is transferred by crossing-over from one chromatid of the Y to one chromatid of the X. Furthermore, closer inspection of XY bivalents in mice not carrying *Sxr* suggests that crossing-over is a normal feature of X–Y pairing, and may be a necessary requirement for the production of functional sperm. Details of the mode of inheritance of *Sxr* are given in Table 1.

Thorough examination of the karyotype of X/Y *Sxr* males and their X/Y non-*Sxr* sibs showed no gross departure from the standard XY karyotype, and it was concluded that the claim for the existence of an X, Y, Y1 system⁶ resulted from misinterpretations of the precocious separation of the Y chromatids at mitotic metaphase when prolonged colcemid treatments are employed (see also refs 2–4), and of the unusual configurations adopted by the X and Y chromosomes in the sex vesicle during meiotic prophase^{8,9}.

Detailed analysis of the mouse Y chromosome is difficult as in metaphase of mitosis it frequently shows uniformly dense staining after G-banding¹⁰, and even if bands are revealed they are often distorted by coiling. In the accepted nomenclature¹⁰ the chromosome is described as being composed of five major bands (Fig. 1a) but the authors included the proviso that examples from some strains might reveal variant banding patterns. We would add that many, if not all, mouse Y chromosomes have a short arm¹¹, and in the case of the elongated examples from the control mice studied here, reveal further bands on their long arms (Fig. 1b).

Examination of the Y chromosome of X/Y males from the *Sxr* stock (Table 1) revealed that whereas X/Y non-*Sxr* animals possessed a normal appearing Y, the Y of X/Y *Sxr* males possessed an additional constricted dark-staining distal band or body (Fig. 1c). In midmetaphase this caused the proximal and distal ends of the Y to be so similarly marked that it became difficult to orientate the chromosome with respect to its centromere. However, the distal body stained in a reverse manner to the centromeric or proximal region of the chromosome, being dark in the early stages and becoming progressively lighter as metaphase proceeds.

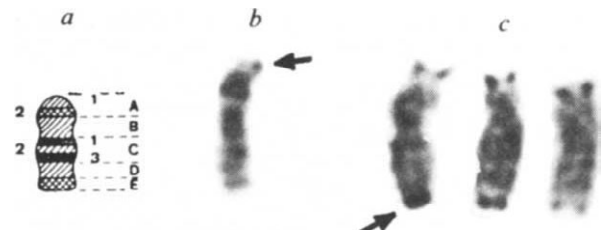


Fig. 1 a, The accepted standard idiogram of the mouse Y chromosome¹². b, G-banded, elongated Y from a bone marrow cell of a normal (control) XY male mouse, showing evidence of a short arm (arrowed) and further bands on the long arm. Control mice are derived from the 101/H, Ju/FaCt and Q (Falconer) strains among which no clear Y chromosomal differences were found. c, G-banded Y chromosomes from bone marrow cells of an X/Y *Sxr* mouse. Left: in early metaphase, showing the additional, heavily stained, distal body (arrowed). Centre: in mid-metaphase, showing the equivalent staining of proximal and distal ends. Right: in late metaphase, showing the fainter staining of the distal body with mitotic progression.

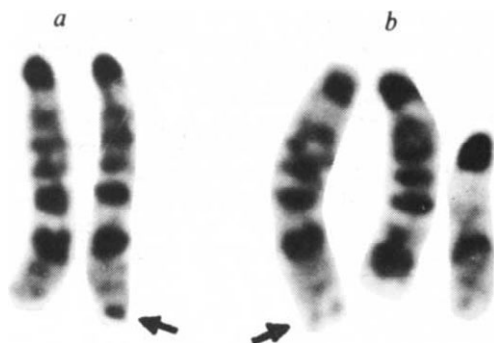


Fig. 2 *a*, G-banded X chromosomes from a bone marrow cell of an X/X *Sxr* male. Left: the normal X; right: the X has an additional, dark-staining, distal body (arrowed). *b*, G-banded X chromosome from a bone marrow cell of an X/X *Sxr* male carrying the T(X;16)16H X-autosome translocation. Left: the paternally-derived normal X showing the distal body (arrowed).

The X chromosome in mitotic cells of X/Y *Sxr* males was observed to conform to the normal banding pattern of this chromosome¹⁰, as were both X chromosomes from their X/X daughters. However, although one X of their X/X *Sxr* sons also looked normal, the other X possessed a dark-staining body at its distal end (Fig. 2*a*). Furthermore, in X/X *Sxr* males carrying the T(X;16)16H X-autosome translocation it was always the paternally-derived, non-translocated X chromosome that carried this body (Fig. 2*b*).

Meiotic studies at diakinesis in X/Y non-*Sxr* males (the X/Y progeny of X/Y *Sxr* mice shown in Table 1) and X/Y (from unrelated stocks) controls showed that in most spermatocytes the X and Y chromatids paired distally with a clear line of separation between them (Fig. 3*a*). In X/Y *Sxr* males the pairing was also distal but one pair of the X and Y chromatids showed more contact by way of two dark staining bodies (Fig. 3*b*), one on the X and the other on the Y. The bodies were always adjoining and not diagonally opposed, which indicated that a chiasmate exchange had occurred whereby the body had crossed over from one chromatid of the Y to one of the X. Confirmatory evidence of the exchange was provided by metaphase II observations of X and Y chromosomes with single chromatids marked by the dark-staining distal body. Another feature of the first meiotic division of X/Y *Sxr* males is a higher than normal frequency of X-Y univalence which, at diakinesis, may vary from 70 to 90%, as against 5 to 10% in the X/Y non-*Sxr* sibs⁷. In many, if not all, spermatocytes with X-Y univalence dark-staining bodies were observed on both chromatids of the Y with none of the X (Fig. 3*c*), suggesting that pairing failure in meiotic prophase gives no opportunity for transfer.

Our observations with *Sxr* mice indicate that *Sxr* represents a Y rearrangement in which part or all of the male determining region has been transferred from a relatively proximal location

on the chromosome to the distal end where, by crossing-over, it can be transferred to the X and thus cause male development in X/X *Sxr* mice (see also refs 2-4). Our mitotic observations with X/Y *Sxr* animals, however, showed no evidence of any chromosomal change in the Y other than the addition of the distal, *Sxr* body. We therefore conclude that the original rearrangement involved either, (1) part of a multiple copy sex-determining region the loss of which from its normal location we cannot distinguish cytologically, or, (2) all or most of this region transferred by excision from one chromatid⁵ and attached onto the distal end of the other. Our meiotic observations clearly show that *Sxr* is inherited in a simple manner after the formation of a regular chiasma and a cross-over taking place between the distal regions of the X and Y. Such a configuration produces four different classes of chromatid, X, X with *Sxr*, Y and Y with *Sxr*, which, when transmitted by the gametes, satisfies the segregation ratios observed in the progeny of X/Y *Sxr* males⁷. Pairing failure gives no opportunity for crossing-over and is manifested by X-Y univalence which, from our metaphase II observations appears to act as a cell lethal.

It may seem surprising in view of the many past cytological studies on *Sxr* mice that the *Sxr* body and its Y to X transfer has gone undetected. In part this may have been due to the fact that the apparent autosomal dominant inheritance focused attention primarily upon the autosomes. However, at the mitotic level, failure may also have been due to the staining vagaries of both the Y and the distal, *Sxr* body; at the meiotic level, to the best of our knowledge, no G-band analyses have previously been attempted on the condensed stages and at the synaptonemal complex (SC) level, any possible differences between X and Y pairing in carrier and non-carrier mice might have been obscured by synaptic adjustment¹².

The question of greatest significance which remains is whether crossing-over between the X and Y chromosomes is limited to X/Y *Sxr* males, or is typical of all mouse XY bivalents. Reappraisal of normal XY bivalents in the light of our *Sxr* findings suggests that with some there is evidence of an asymmetry caused by a greater continuity of adjoining pairs of X and Y chromatids (Fig. 3*d*), and this could be taken as evidence of a residual chiasma.

However, this behaviour was only observed in a minority of cells. The apparent inconsistency may be resolved by examining the available evidence for the cytologically visible amount of chromatin expected to be exchanged by crossing-over. Studies at the SC level¹³ suggests that there is considerable pairing segment between the X and Y chromosomes and the possibility of a fairly large exchange. However, this can be excluded for the following reasons: (1) a cytological reassessment of the large paracentric inversion of the X chromosome¹⁴, In(X)1H, identifies the distal break-point as being in band F3, or even F4, which is the terminal band, and despite this, the sex-bivalent

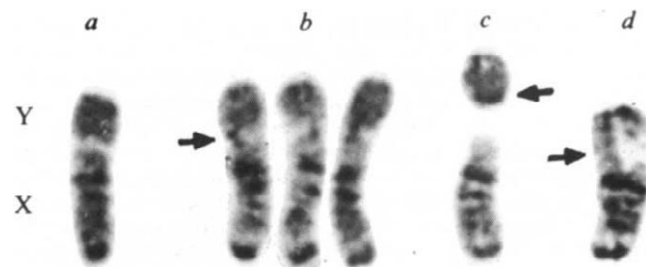


Fig. 3 G-banded X and Y chromosomes at diakinesis. *a*, From a control X/Y male showing the distal pairing of X and Y with a clear line of separation. *b*, Three examples from an X/Y *Sxr* male showing a greater continuity of one pair of adjoining X and Y chromatids by way of two dark staining bodies (arrowed). *c*, Univalent X and Y from an X/Y *Sxr* male showing the retention of the dark staining bodies on both chromatids of the Y (arrowed). *d*, Sex-bivalent from an X/Y non-*Sxr* male showing the greater continuity of an adjoining pair of X and Y chromatids (arrowed) suggesting earlier chiasma formation.

Table 1 The inheritance of *Sxr*

Cross	X/X ♀	×	X/Y <i>Sxr</i> ♂
Progeny	X/X ♀		X/Y ♂
	X/X <i>Sxr</i> ♂		X/Y <i>Sxr</i> ♂

The sex-reversed (*Sxr*) factor causes chromosomal females to develop as phenotypically normal, but sterile males⁵. Transmission is through X/Y carrier males and, hence, all X/Y males, whether carrying *Sxr* or not, normally have the same Y chromosome. The mode of inheritance appears to be that of an autosomal dominant, but all attempts to locate *Sxr* on an autosome have failed⁷. A location in an homologous region of the X and Y chromosomes could give the same pattern of inheritance, however². The observed segregation ratios accord well with expectation, the X/X and X/Y classes of progeny usually being distinguished with the aid of an X chromosomal marker gene. Stock maintenance is by crossing X/Y *Sxr* males × F₁ hybrid (C3H/HeH ♀ × 101/H ♂) females.

in male carriers shows a near-normal frequency of X-Y pairing at diakinesis, (2) cream (*Crn*), the most distal gene on the mouse X chromosome, lies outside this inversion¹⁶ and yet in male mice with or without the inversion it has never been found to cross-over onto the Y (Lyon, M. F., personal communication). The amount of chromatin visibly available for reciprocal exchange is therefore minimal and the chiasma between the X and Y must be near-terminal, as has previously been suggested from SC studies¹⁶. It is therefore not surprising that the chiasma is displayed at diakinesis as 'terminalized' and the bivalent as symmetrical in the majority of primary spermatocytes of normal males which do not have the *Sxr* body to serve as a visible cross-over marker. We conclude that in all male mice a small reciprocal exchange between the X and Y is a regular and necessary requirement of meiosis for the production of functional spermatozoa.

Received 15 July; accepted 8 October 1982.

- Crewe, F. A. E. & Koller, P. C. *J. Genet.* **26**, 359-383 (1932).
- Burgoyne, P. B. *Hum. Genet.* (in the press).
- Eicher, E. M. in *Prospects for Sexing Mammalian Sperm* (eds Amann, R. T. & Seidell, G. E.) (Colorado University Press, Boulder, 1982).
- Hansmann, I. *Cell* **30**, 331-332 (1982).
- Cattanach, B. M., Pollard, C. E. & Hawkes, S. G. *Cytogenetics* **10**, 318-337 (1971).
- Singh, L. & Jones, K. W. *Cell* **28**, 205-216 (1982).
- Lyon, M. F., Cattanach, B. M. & Charlton, H. M. in *Mechanisms of Sex Differentiation in Mammals* (eds Austin, C. R. & Edwards, R. G.) 329-386 (Academic, New York, 1981).
- Chandley, A. C. & Fletcher, J. M. *Chromosoma* **81**, 9-17 (1980).
- Evans, E. P., Burtenshaw, M. D. & Brown, B. B. *Chromosoma* **81**, 19-26 (1980).
- Nesbitt, M. N. & Francke, V. *Chromosoma* **41**, 145-158 (1973).
- Ford, C. E. *Transplantation* **4**, 333-335 (1966).
- Moses, M. J. *et al. J. Cell Biol.* **79**, 123a (1978).
- Tres, L. L. *J. Cell Sci.* **25**, 1-15 (1977).
- Evans, E. P. & Phillips, R. J. *S. Nature* **256**, 40-41 (1975).
- Fisher, G. & Lyon, M. F. *Mouse News Lett.* **64**, 58 (1981).
- Solari, A. G. *Chromosoma* **29**, 217-236 (1970).

Male, female and intersex development in mice of identical chromosome constitution

B. M. Cattanach*, E. P. Evans†, M. D. Burtenshaw† & J. Barlow*

*MRC Radiobiology Unit, Harwell, Didcot, Oxfordshire OX11 0RD, UK

†Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE, UK

In the preceding paper¹, cytological evidence is presented to show that in 'sex-reversed' XX male mice² the sex-reversing factor, *Sxr*, is carried on a dark-staining body located on the distal end of the paternal X chromosome. The X-chromosomal location of *Sxr* raises the possibility that it may be subject to the X chromosome inactivation process³. However, the regular segregation of *Sxr* and the absence of any indication that intersexes occur with a higher than normal frequency in *Sxr* stocks⁴ suggests that either inactivation of *Sxr* does not occur, or, if it does, that the proportion of cells with the *Sxr*-bearing X genetically active is sufficient to ensure normal testicular differentiation. Here we present information on the sexual development of chromosomally XX mice carrying *Sxr* in which the *Sxr*-bearing X chromosome is held in the inactive condition in most if not all somatic cells by use of the T(X;16)16H (T16H) X-autosome translocated chromosome⁵⁻¹⁰. Mice of this genotype were found to develop as normal but sterile males, normal fertile females or intersexes.

That T16H/X *Sxr* animals can develop as males was indicated by two observations. First, the sex ratio among the T16H progeny of the cross of T16H/X females with X/Y *Sxr* males (36:87, Table 1) was not significantly different ($\chi^2 = 0.98$) from the 1:3 (females:males) ratio normally found in *Sxr* stocks⁴,

Table 1 Breeding of T16H/X *Sxr* mice

Parental genotypes	T16H++/++ <i>Blo</i> ♀ × <i>Ta</i> +/ <i>Y Sxr</i> ♂	
T16H progeny	Expected sex	No. observed*
T16H++/++ <i>Ta</i> +	♀	36 ♀
T16H++/++ <i>Ta</i> + <i>Sxr</i>	♀ or ♂?	
T16H++/ <i>Y</i>	♂	87 ♂
T16H++/ <i>Y Sxr</i>	♂	
Non-T16H progeny		
++ <i>Blo</i> /++ <i>Ta</i> +	♀	36 ♀
++ <i>Blo</i> /++ <i>Ta</i> + <i>Sxr</i>	♂	34 ♂
++ <i>Blo</i> / <i>Y</i>	♂	57 ♂
++ <i>Blo</i> / <i>Y Sxr</i>	♂	

* Based on external sexual phenotype.

and also evident among the non-T16H progeny of the cross 36:91, Table 1). Second, examination of the testes of 61 of the 87 males carrying T16H revealed that 17 lacked germ cells and were thus genetically T16H/X *Sxr*², and that the remaining 44 possessed germ cells, and were therefore genetically either T16H/Y or T16H/Y *Sxr* (Table 1). Cytological confirmation of the chromosome constitution was obtained in several animals. The T16H/X:T16H/Y plus T16H/Y *Sxr* ratio (17:44) agrees well with the 1:2 expectation ($\chi^2 = 0.59$), and provides little reason to suspect that T16H/X *Sxr* animals might develop other than as males. The data at this point therefore suggest that *Sxr* is not inactivated when the X on which it is located is inactivated.

Investigation of the T16H females derived from the cross provided reason to modify this conclusion, however. On being test-mated with a normal male, one female produced, among a total of 21 progeny, five X/X *Sxr* sons and two T16H/X *Sxr* sons (Table 2). Clearly, this female must have carried *Sxr*, and this was confirmed by the cytological observation of the *Sxr* body on its normal X chromosome. The occurrence of the T16H/X *Sxr* sons, and also of two non-T16H daughters (Table 2), incidentally provided evidence of meiotic recombination between the T16H breakpoint and *Sxr* in the mother, transferring *Sxr* from the normal X to the T16H X chromosome. Eight other fertile females have now been examined cytologically, and two carried *Sxr*. Thus, among a total of 20 T16H/X *Sxr* animals so far examined, 3 have developed as females and 17 as males. The female development of some T16H/X *Sxr* animals indicates that *Sxr* must be subject to the X-inactivation process in some cells.

Further evidence that *Sxr* can be subject to X-inactivation was obtained from more detailed examination of the T16H/X *Sxr* males. Although all 17 appeared unambiguously male, internally as well as externally, four possessed testes which displayed abnormalities greater than those typically found in X/X *Sxr* (that is, non-T16H) males: their testes tended to be smaller, their tubules appeared more difficult to tease apart than the equivalent sterile tubules of X/X *Sxr* males, and, most strikingly, they were partially covered by blood clots which extended to the epididymis and proximal regions of the vas deferens. One testis from each of these males has been examined histologically and in each one a minor (5-10%) ovarian component was identified. The remaining 13 males possessed histologically normal testes. The occurrence of intersex T16H/X *Sxr* mice is consistent with McLaren and Monk's¹¹ discovery of three hermaphrodites, with a testis on one side and an ovary on the other, among 10 T16H/X *Sxr* mice.

Two of the three T16H/X *Sxr* females identified have also been examined in detail. In both, the ovaries and reproductive organs were macroscopically normal and the structure of the ovaries was histologically normal. Clearly, therefore, T16H/X *Sxr* mice may show a range of phenotypes from normal male, through intersex, to normal female. An incidental point is that the fertility of the T16H/X *Sxr* females shows that the presence

Table 2 Progeny of a fertile T16H/X *Sxr* female

Cross: T16H++/+Ta+ <i>Sxr</i> ♀ × ++Blo/Y ♂			
XX progeny		XY progeny	
Genotype	No.	Genotype	No.
T16H++/+Blo ♀	1	T16H++/Y ♂	?
T16H++ <i>Sxr</i> /++Blo ♂*	2	T16H++ <i>Sxr</i> /Y ♂*	?
+Ta+ <i>Sxr</i> /++Blo ♂	5	+Ta+ <i>Sxr</i> /Y ♂†	1
+Ta+/++Blo ♀*	2	+Ta+/Y ♂*	?
			9
			2

* T16H-*Sxr* recombinant classes. Recombination frequency 36.5% (4/11).

† Novel XY genotype in which *Sxr* must be carried on the X rather than on the Y.

in the oocytes of the male-determining (*Sxr*) DNA sequences in a form which is presumably genetically active (as both X chromosomes are reactivated at the onset of meiosis^{12,13}) does not interfere with their ability to form functional oocytes. This observation is consistent with the finding¹⁴ that XY germ cells could form functional oocytes in a female XX↔XY mouse chimaera.

The gene markers used in the T16H cross (Table 1) clearly showed that the presence of *Sxr* did not modify the non-random X-chromosome activity in the T16H heterozygote. The range of sexual development in animals of a single (T16H/X *Sxr*) genotype therefore suggests that the primary cause is a variable inactivation of *Sxr* in the inactivated *Sxr*-bearing X, generating *Sxr*-active and *Sxr*-inactive populations of cells. As such this may represent a further example of position effect variegation in mammals¹⁵, so far only observed with autosomal loci in certain X-autosome translocations^{16,17}. Clearly, secondary factors also must operate in T16H/X *Sxr* mice to influence the final sexual phenotype, and these would include frequency and distribution of the two cell types, the ability of the *Sxr*-active cells to induce *Sxr*-inactive cells to form testicular tissue, perhaps through H-Y antigen coating (as postulated to occur in XX↔XY chimaeras¹⁸) and also the number of cells in the fetus that form the primordial gonad.

The fact that male development occurs at all in T16H/X *Sxr* mice where the *Sxr*-bearing X is inactive in virtually all somatic cells, indicates that *Sxr* must remain active, although on an inactive X, in a high proportion of cells. Thus, with random X-inactivation, the proportion of cells with *Sxr* inactive must be very low, and it is therefore likely that X/X *Sxr* animals seldom, if ever, develop as females or intersexes due to the presence of *Sxr*. The possibility is, however, being investigated in X/X *Sxr* animals which, because of X chromosome controlling element (*Xce*) allele heterozygosity, have the *Sxr*-bearing X active in only about 30% of their somatic cells¹⁹.

Fertile females produced by inactivation of an X chromosome of 'sex-reversed' mice

Anne McLaren & Marilyn Monk

MRC Mammalian Development Unit, Wolfson House (University College London), 4 Stephenson Way, London NW1 2HE, UK

Sex-reversed (*Sxr*) is a dominant mutation that confers maleness on X/X *Sxr* mice¹. Although its mode of inheritance is typically autosomal, all attempts to map *Sxr* to an autosomal chromosome have failed². P. Burgoyne (personal communication) suggested that if *Sxr* were located at the end of the sex chromosomes, distal to a postulated obligatory cross-over in the X-Y pairing segment, it would show such an apparent autosomal pattern of inheritance. To test whether *Sxr* is located on one of the X chromosomes, and is perhaps therefore subject to X-chromosome inactivation, we have examined mice heterozygous for an X-autosome translocation, T(X; 16)H (ref. 3), such that the normal X chromosome postulated to carry *Sxr* is preferentially inactive. Independently, Singh and Jones⁴ have now shown by *in situ* hybridization that *Sxr* consists of a duplicated section of Y chromosome material that is indeed translocated to the distal terminus of one X chromatid during male meiosis^{5,6}. Our results show that when the X chromosome carrying *Sxr* is preferentially inactivated, fertile T(X; 16)H/X *Sxr* females can be produced, owing to an associated inactivation of the male-determining *Sxr* sequences.

Females carrying T(X; 16)H, hereafter referred to as T16H, and heterozygous for the X-linked enzyme marker *Pgk-1*^{7,8}, were mated to males carrying *Sxr* and an X-chromosome marker (Table 1). The genotypes of the progeny were deduced as indicated in Table 1. PGK-1 phenotype was determined in order to classify the maternally inherited X chromosome (T16H *Pgk-1*^b or X *Pgk-1*^a). Male and female *Sxr*-carrying progeny were identified by the presence in their progeny of males with variegated coats after test mating with partners carrying an X-linked coat pattern marker. Sterile males were analysed by a combination of testis weight, testis histology and sex chromosome constitution, to distinguish between T16H/Y males in which spermatogenesis is arrested, and X/X *Sxr* males in which it is absent. The results are shown in Table 2. Progeny inheriting the normal X chromosome from the mother served as controls for those inheriting the T16H chromosome. A deficiency of *Sxr* carriers was observed among the progeny regardless of which X chromosome they inherited from the mother. All the female progeny were fertile. All the X/X *Sxr* progeny (13/13) were sterile males, showing typical X/X *Sxr* testicular histology¹. However, of the T16H/X *Sxr* progeny, five were sterile males, two developed as fertile females, and three proved to be hermaphrodites, with a testis on one side and an ovotestis on the other. The ovotestes were very small, and contained oocytes, sometimes in well developed follicles, within abnormal testicular tubules.

Identification of the two females carrying *Sxr* was obtained by progeny-testing. Table 3 shows the data for one of the females, which was mated first to a Gs/Y and then to a Ta/Y male. The female produced two phenotypically striped X/X *Sxr* sons, and hence was carrying *Sxr*. The overall sex ratio (23♂:11♀) differs significantly from 1:1 ($P < 0.05$), but not from the 3:1 ratio expected from an *Sxr*-carrying female. Moreover, amongst the non-variegated male young five T16H *Sxr*/X males (non-variegated owing to non-random X-chromosome expression) were identified, indicating that *Sxr*, initially present on the *Pgk-1*^a X chromosome, may be transferred to the T16H chromosome during oogenesis. The proportion of XX progeny inheriting *Sxr* from the mother (7/18) is consistent with the 50% expected from an *Sxr* heterozygote.

The finding that *Sxr* sometimes fails to be expressed has been confirmed by Cattanaach and his colleagues⁹, who have found

Received 15 July; accepted 8 October 1982.

- Evans, E. P., Burtenshaw, M. & Cattanaach, B. M. *Nature* **300**, 443-445 (1982).
- Cattanaach, B. M., Pollard, C. E. & Hawkes, S. G. *Cytogenetics* **10**, 318-337 (1971).
- Lyon, M. F. *Nature* **190**, 372-373 (1961).
- Lyon, M. F., Cattanaach, B. M. & Charlton, H. in *Mechanisms of Sex Determination in Animals and Man* (eds Austin, C. R. & Edwards, R. G.) 329-386 (Academic, New York, 1981).
- Lyon, M. F. *et al.* *Cytogenetics* **3**, 306-323 (1964).
- Eicher, E. M. in *Prospects for Sexing Mammalian Sperm* (eds Amann, C. R. & Seidell, G. E.) (Colorado University Press, 1982).
- Ohno, S. & Lyon, M. F. *Chromosoma* **16**, 90-100 (1965).
- Russell, L. B. & Cachiero, N. L. A. in *Genetic Mosaics and Chimeras in Mammals* (Basic Life Sciences, Vol. 12 (ed. Russell, L. B.) 393-416 (Plenum, New York, 1978).
- Tagaki, N. *Chromosoma* **81**, 439-459 (1980).
- Disteche, C. M., Eicher, E. M. & Latt, S. A. *Expl Cell Res.* **133**, 357-362 (1981).
- McLaren, A. & Monk, M. *Nature* **300**, 446-448 (1982).
- Gartler, S. M., Andina, R. & Gant, N. *Expl Cell Res.* **91**, 454-457 (1975).
- Johnston, P. G. *Genet. Res.* **37**, 312-322 (1981).
- Evans, E. P., Ford, C. E. & Lyon, M. F. *Nature* **267**, 430-431 (1977).
- Cattanaach, B. M. *Genet. Res.* **23**, 291-306 (1974).
- Cattanaach, B. M. *Rev. Genet.* **9**, 1-18 (1975).
- Eicher, E. M. *Adv. Genet.* **15**, 175-259 (1970).
- Ohno, S. *Cell* **7**, 315-321 (1976).
- Johnston, P. & Cattanaach, B. M. *Genet. Res.* **37**, 151-160 (1981).

Table 1 Characters scored in the genotype analysis of the progeny of the parental cross T16H *Pgk-1*^b/+ *Pgk-1*^a ♀ × *Pgk-1*^a/Y *Sxr* ♂

Phenotypic sex	PGK-1 type*	Fertility†	Analysis of sterile males‡			Appearance of <i>Sxr</i> in F ₂ progeny§		Genotype
			Testis weight (mg)	Spermatogenesis (histology)	Karyotype	Variegated XX <i>Sxr</i> males	Non-variegated T16H/X <i>Sxr</i> males	
♀	A	Fertile	—	—	—	Absent	—	+ <i>Pgk-1</i> ^a /+ <i>Pgk-1</i> ^a
♀	B	Fertile	—	—	—	Absent	Absent	T16H <i>Pgk-1</i> ^b /+ <i>Pgk-1</i> ^a
♀	B	Fertile	—	—	—	Present	Present	T16H <i>Pgk-1</i> ^b /+ <i>Pgk-1</i> ^a <i>Sxr</i>
♂	B	Sterile	<25	Absent	XX	—	—	
♂	B	Sterile	10–50	Arrested	XY	—	—	T16H <i>Pgk-1</i> ^b /Y, or Y <i>Sxr</i>
♂	A	Fertile	—	—	—	Absent	—	+ <i>Pgk-1</i> ^a /Y
♂	A	Fertile	—	—	—	Present	—	+ <i>Pgk-1</i> ^a /Y <i>Sxr</i>
♂	A	Sterile	<25	Absent	XX	—	—	+ <i>Pgk-1</i> ^a /+ <i>Pgk-1</i> ^a <i>Sxr</i>

The female mice in the parental cross were derived from crosses between a stock carrying T16H, obtained originally from MRC Radiobiology Unit, Harwell, and a stock carrying *Pgk-1*^a, a variant allele at the X-linked locus coding for the monomeric enzyme PGK-1 (phosphoglycerate kinase 1; EC2.7.2.3), obtained originally from John West, Zoology Department, Oxford. The male mice in the parental cross carried either *Pgk-1*^a or one of the X-linked coat-pattern markers *Ta* (Tabby), *Hq* (Harlequin) and *Gs* (Greasy), obtained originally from the Institute of Animal Genetics, Edinburgh. The table outlines the strategy used for classifying genotypically the progeny derived from *Pgk-1*^a/Y *Sxr* males; a similar strategy was used for the progeny of *Ta*/Y, *Gs*/Y and *Hq*/Y males. The strategy also allowed the identification of recombinant progeny; these have not been included in the table.

* PGK-1 (A, B or AB) was determined by gel electrophoresis²⁸ of blood samples taken from the retro-orbital sinus.

† To test fertility, each of the male progeny was mated to two females carrying *Ta*, *Gs* or *Hq*, and each of the female progeny was mated to a male carrying one of the same three markers. Inspection of the coat pattern of the young of fertile animals allowed *Sxr* carriers to be identified (see §) and XX to be distinguished from T16H/X females (the latter have some non-variegated daughters).

‡ Sterile males were killed and their testes weighed. A plot of mean testis weight against body weight allowed X/Y to be distinguished reliably from T16H/Y males. Non X/Y testes weighing at least 25 mg were classified as T16H/Y; below 25 mg there was a region of overlap between the smaller T16H/Y and the larger X/X *Sxr* testes. These testes were therefore analysed histologically (spermatogenesis is absent in X/X *Sxr* and arrested in T16H/Y males); in most cases the mice were also karyotyped from bone marrow preparations. We could not distinguish T16H/Y *Sxr* from T16H/Y sterile males.

§ Fertile males and females carrying *Sxr* were identified by the presence of variegated (striped) males (X/X *Sxr*) among their progeny (for matings, see †). In the case of PGK-1B females (putatively T16H), non-variegated male progeny were also screened by the analysis described above ‡, since the T16H/X *Sxr* sons, although carrying a *Ta*, *Gs* or *Hq* gene on their normal chromosome, would not be variegated.

another fertile T16H/X *Sxr* female, shown to be an *Sxr* carrier by progeny testing, as well as two other T16H/X *Sxr* females identified cytologically. The occurrence of such females among T16H/X *Sxr* but not X/X *Sxr* individuals (Table 2) suggests strongly that the failure of expression is due to inactivation of the *Sxr* male-determining region as a consequence of its attachment to the normal X chromosome, preferentially inactive in the presence of T16H.

The observation that some T16H/X *Sxr* individuals develop as females, some as males, and some as hermaphrodites, is most simply explained by assuming that the *Sxr* male-determining region is inactivated in some cells but not others. Inactivation may spread to a variable extent from the inactive X chromatin into the attached *Sxr* fragment, in an analogous fashion to position effect variegation for autosomal genes translocated onto the inactive X chromosome¹⁰. We assume that the extent of spreading leading to active or inactive *Sxr* status is a heritable characteristic determined at the time of X inactivation. The sex of each individual will then depend on the proportion of cells expressing *Sxr* in the gonad primordium.

Studies on aggregation chimaeras suggest that a certain critical proportion of XY male-determining cells is required for a mixed gonad to develop as male. Of 31 adult XX↔XY chimaeras reported in 5 different strain combinations^{11–15}, 21 were male, 7 female and 3 hermaphrodite. Since the relative proportions of the two cell components appear to show a flat distribution from 0 to 100% in aggregation chimaeras¹⁶, this would indicate (ignoring variation from one strain combination to another) that the critical threshold above which development of a normal male gonad occurs is about 32% (10/31) XY cells. In the T16H/X *Sxr* animals that develop as females we propose that this critical proportion of cells with *Sxr* active was not reached. The hermaphrodites would be explained by an *Sxr* mosaicism in the region of threshold proportion, whereas the males of the same genotype would have proportions of cells with *Sxr* active above the critical levels. The relative proportions

will vary from individual to individual, and even between the two gonads within a single individual, leading to unilateral hermaphroditism.

If the extent of spreading is linked to the primary X-inactivation event, the distribution of *Sxr* mosaicism among individuals will reflect the small number of cells present when inactivation occurs. The number of progenitor cells present in the embryo at this time has been estimated to be between 13 and 64¹⁷. Assuming a progenitor cell pool of at least 13 cells, and a requirement of about 30% gonadal cells with *Sxr* active over the period of testis determination for normal male development, a calculation based on binomial expectations suggests that inactivation would need to spread into *Sxr* in more than half the inactive X chromosomes to account for the observed number of hermaphrodites and females among T16H/X *Sxr* individuals.

Our observation of adult T16H/X *Sxr* hermaphrodites is consistent with earlier reports that X/X *Sxr* germ cells occasionally entered meiosis before birth¹⁸, sometimes in association

Table 2 Progeny of T16H *Pgk-1*^b/+ *Pgk-1*^a ♀ × X/Y *Sxr* ♂ matings

	Chromosome from male				Total
	X	X <i>Sxr</i>	Y	Y <i>Sxr</i>	
Chromosome from female (X)					
T16H <i>Pgk-1</i> ^b	18♀	2♀ 5♂ 3♀	36♂		64
+ <i>Pgk-1</i> ^a	23♀	13♂	24♂	11♂	71
Recombinant	2♀	2♂	3♂		7
Total	43	25	74		142

The progeny of the different types of *Sxr* ♂ did not differ significantly from one another and have been combined. The eight non-recombinant classes are expected to be produced in equal proportions.

Table 3 Progeny of T16H *Pgk-1^b* / + *Pgk-1^a* *Sxr* ♂

		From ♂*	
		X	Y
From ♀	T16H <i>Pgk-1^b</i> <i>Sxr</i>	5♂	10♂
	T16H <i>Pgk-1^b</i>	5♀	
	X <i>Pgk-1^a</i> <i>Sxr</i>	2♂†	5♂
	X <i>Pgk-1^a</i>	6♀	
	Recombinant	—	1♂

* A *Pgk-1^b* *Gs/Y* male was used to sire the first three litters, and a *Pgk-1^b* *Ta/Y* male for subsequent litters. Both males were extensively tested with normal females to ensure that they were not carrying *Sxr*.

† Phenotypically variegated coats.

with disturbed seminiferous tubule development¹⁹. A small proportion of the meiotic germ cells survived birth, to give rise to growing oocytes within the tubules, but no oocytes were seen in 30 adult X/X *Sxr* testes²⁰. On their own, these earlier observations of ovotestes in young X/X *Sxr* animals could reflect merely some inadequacy in the masculinizing influence of *Sxr*, and/or the presence of two X chromosomes in the germ cells; but in the light of the present results on T16H/X *Sxr* animals, the earlier findings too can be taken to indicate that *Sxr* may be subject to X inactivation, even in X/X *Sxr* individuals⁶. The proportion of cells with *Sxr* inactivated by spreading will of course be lower in X/X *Sxr* than in T16H/X *Sxr* individuals, since only about half the *Sxr*-carrying X chromosomes will be inactive in the former, but all or almost all in the latter. An X/X *Sxr* individual in which sufficiently few cells expressed the male-determining *Sxr* sequences to allow the development of a fertile female would therefore be very rare indeed.

Two recently formulated sex chromosome models include the prediction that X-chromosome material in an embryo will be protected from inactivation if it has been paired during the preceding male meiosis^{5,21}. Burgoyne⁵ has accordingly suggested an explanation for the variable inactivation of *Sxr* when it is attached to an inactive X chromosome, namely that the expression of *Sxr* may be associated with a folding back and pairing of the *Sxr* region with the homologous testis-determining region in the same Y chromosome during the preceding male meiosis. If such pairing has occurred during meiosis, the *Sxr* region would be protected from subsequent inactivation in the embryo. To account for the observed hermaphroditism, one could further postulate that multiple male-determining sequences exist and that partial intra-chromosomal Y pairing may be followed by inactivation of a variable number of these sequences.

We did not know whether a T16H/X *Sxr* female, if she existed, would be fertile or sterile. The inactive X chromosome in female germ cells of the mouse is known to be reactivated at or shortly before entry of the germ cells into meiotic prophase^{7,22–24}. In oocytes, therefore, both X chromosomes are expressed. The finding that female *Sxr* carriers can be fully fertile implies that the presence in female germ cells of male-determining DNA sequences on an active X chromosome from the time of entry into meiosis onwards does not impede the progress of the germ cells through oogenesis. This conclusion is consistent with a report¹³ of a fertile female XX↔XY mouse chimaera that produced a functional XY egg, giving rise after fertilization to an XXY embryo. In another such chimaera, a single XY oocyte in first meiotic metaphase was identified cytologically²⁵. A contrary conclusion, namely that the presence of *Sxr* prevents XX germ cells from undergoing oogenesis, was reached by Gordon²⁶ on the basis of a series of experimental chimaeras thought to have an X/X *Sxr* component. The evidence rests on a single female chimaera that produced germ cells from the non-*Sxr* component only; however, as the author himself points out, this could have been brought about by chance alone. Probably the chimaeras that involved an X/X *Sxr* component developed as phenotypic males.

In a recent study²⁷, *Drosophila* homozygous for mutant autosomal loci involved in sex determination developed sterile gonads devoid of germ cells. However, when the germ cells were introduced by pole cell transplantation into a normal embryo of the appropriate sex, they gave rise to normal functional gametes, suggesting that the mutant sex-determining loci did not affect development in the germ cells themselves. Our finding that the presence of *Sxr* in female germ cells does not interfere with oogenesis suggests that the same situation prevails in mice.

Received 20 July; accepted 30 September 1982.

- Cattanach, B. M., Pollard, C. E. & Hawkes, S. G. *Cytogenetics* **10**, 318–337 (1971).
- Lyon, M. F., Cattanach, B. M. & Charlton, H. M. in *Mechanisms of Sex Differentiation in Animals and Man* (eds Austin, C. R. & Edwards, R. G.) 329–386 (Academic, New York, 1981).
- Lyon, M. F., Searle, A. G., Ford, C. E. & Ohno, S. *Cytogenetics* **3**, 306–323 (1964).
- Singh, L. & Jones, K. W. *Cell* **28**, 208–216 (1982).
- Burgoyne, P. *Hum. Genet.* **61**, 85–90 (1982).
- Eicher, E. M. in *Prospects for Sexing Mammalian Sperm* (eds Amann, R. P. & Seidel, G. E.) (Colorado Associated University Press, Colorado, in the press).
- Johnston, P. G. *Genet. Res.* **37**, 317–322 (1981).
- McMahon, A. & Monk, M. *Genet. Res.* (in the press).
- Cattanach, B. M., Evans, E. P., Burtenshaw, M. D. & Barlow, J. *Nature* **300**, 445–446 (1982).
- Cattanach, B. M. *Genet. Res.* **23**, 291–306 (1974).
- Mytowska, E. T. & Tarkowski, A. K. *J. Embryol. exp. Morph.* **20**, 33–52 (1968).
- Millet, R. G., Mukherjee, B. B. & Whitten, W. K. *Can. J. Genet. Cytol.* **14**, 933–941 (1972).
- Ford, C. E., Evans, E. P., Burtenshaw, M. D., Clegg, H. M., Tuffrey, M. & Barnes, R. D. *Proc. R. Soc. B* **190**, 187–197 (1975).
- McLaren, A. J. *Embryol. exp. Morph.* **33**, 205–216 (1975).
- Gearhart, J. D. & Oster-Granite, N. L. *Biol. Reprod. Suppl.* **24**, 713–722 (1981).
- Falconer, D. S. & Avery, P. J. *J. Embryol. exp. Morph.* **43**, 195–219 (1978).
- Nesbitt, M. N. *Dev. Biol.* **26**, 252–263 (1971).
- McLaren, A. J. *Reprod. Fert.* **61**, 461–467 (1981).
- Mittwoch, U. & Buehr, M. *Differentiation* **1**, 219–224 (1973).
- McLaren, A. *Nature* **283**, 688–689 (1980).
- Polani, P. *Hum. Genet.* **60**, 207–211 (1982).
- McLaren, A. & Monk, M. *J. Reprod. Fert.* **63**, 533–537 (1981).
- Gartler, S. M., Rivest, M. & Cole, R. E. *Cytogenet. & Cell Genet.* **28**, 203–207 (1980).
- Kratzer, P. G. & Chapman, V. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3093–3097 (1981).
- Evans, E. P., Ford, C. E. & Lyon, M. F. *Nature* **267**, 430–431 (1977).
- Gordon, J. J. *exp. Zool.* **367**–374 (1976).
- Schüpbach, T. *Dev. Biol.* **89**, 117–127 (1982).
- Bücher, T., Bender, W., Fundele, R., Hofner, H. & Linke, I. *FEBS Lett.* **115**, 319–324 (1981).

Synchronized bursting of CA1 hippocampal pyramidal cells in the absence of synaptic transmission

J. G. R. Jefferys* & H. L. Haas

Department of Neurosurgery, University of Zurich, 8091 Switzerland

The synchronization of neuronal firing, seen at its most dramatic in the epilepsies, has generally been attributed to synaptic interactions^{1–5}. We have now discovered a rhythmic spontaneous bursting activity produced by non-synaptic mechanisms. It develops in rat hippocampal slices after chemical synaptic transmission has been blocked by incubation in low Ca^{2+} , increased Mg^{2+} solutions, and persists with almost clockwork regularity for several hours. We report here that spontaneous firing of all the neurones recorded in the slice increased, consistent with the known effects of Ca^{2+} on membrane properties and synaptic transmission^{6–10}, but the synchronous 'field bursts', and presumably their underlying mechanisms, were restricted to the population of pyramidal neurones in the hippocampal CA1 region. Thus, these low Ca^{2+} field bursts are different from the Ca^{2+} -dependent synchronous bursts induced in slices by penicillin which originate in the population of pyramidal cells of the CA3 region^{1–3}.

* Permanent address: The Sobell Department of Neurophysiology, Institute of Neurology, Queen Square, London WC1N 3BG, UK.

While the range of conditions that can elicit these field bursts remains to be established, it is particularly interesting that Ca^{2+} *in vivo* can drop to low levels, and starts to fall before the onset of seizures induced by pentylenetetrazol¹¹. The precise roles in the field burst of non-synaptic mechanisms such as ephaptic interactions and K^+ movements have yet to be determined. The interburst interval, but not the burst size or duration, has proved particularly sensitive to manipulation of the Na^+ - K^+ pump and also to a variety of drugs and transmitters in nano- to micromolar concentrations.

The methods of cutting and maintaining brain slices in our laboratories have been described previously¹²⁻¹⁴ (see Fig. 1 legend). Although incubation of hippocampal slices in low (0.2 mM) Ca^{2+} , high (up to 5 mM) Mg^{2+} solution blocked synaptic responses completely after about 5–10 min, the bursting behaviour described here developed after much more prolonged exposure, typically 1–2 h, and persisted for over 12 h. It was reversed on returning to 2 mM Ca^{2+} solution. At first the bursts consisted of small groups of rather irregular population discharges which grew into mature 'field bursts' (see Fig. 1). These bursts lasted ~4 s and consisted of a large number of population spikes, each of which represented the synchronous discharge of many pyramidal cells, superimposed on a negative potential shift. The field bursts recurred with remarkable regularity, in this case every 21.6 s with a standard deviation (s.d.) of 1.6 s, with little variation in amplitude (mean, excluding spikes, 5.41 mV, s.d. 0.08 mV) or duration (mean 4.32 s, s.d. 0.34 s) during 45 min. Overall interburst intervals ranged from 15 s to several min, the amplitude up to 12 mV, and the burst duration from 0.5 to 20 s. The field bursts occurred synchronously throughout the CA1 in each of 48 investigated slices (12 rats). The two recording sites in Fig. 1a, b were 1 mm apart in the CA1 pyramidal layer. The bursts were closely synchronized at these and all other recording sites. No single site consistently discharged in advance of the rest of CA1, thus there is no evidence for a group of 'pacemaker' pyramidal cells. Incubation in the low Ca^{2+} solutions also changed the antidromic response evoked in the CA1 pyramidal layer by stimulation of the alveus from a single population spike to a burst of 5–20 population spikes. These brief evoked bursts could often trigger a full field burst. The other pyramidal cells (CA3 and 4) and the granule cells in the slice exhibited no spontaneous or evoked periodic behaviour. Therefore, the present phenomenon differs from the bursting induced by penicillin and bicuculline which is Ca^{2+} -dependent, always involves CA3, and consists of briefer bursts¹⁻³.

To explore the cellular basis of the low Ca^{2+} field bursts, we impaled neurones while the slices were bathed in normal Ca^{2+} solutions and held them during the change to low Ca^{2+} and the subsequent development of the bursts. The technical difficulty of this procedure has restricted the number of observations to four, but the comparisons are particularly reliable as they avoid problems of interneuronal variation and sampling errors. Results consistent with these prolonged recordings were obtained from a further 19 cells impaled while in low Ca^{2+} . Exposure to low Ca^{2+} increased the input resistance by 25% if the resting potential were held to its initial value by current injection. Otherwise the slowly developing ~20 mV depolarization was associated with an unchanged or reduced resistance. There was no detectable change in the threshold voltage for action potentials.

The intracellular data suggest that both the slow depolarization of the resting potential and the increased spontaneous discharge (from ~1 s⁻¹ to 10 s⁻¹) following the change to low Ca^{2+} solution resulted from the block of inhibition by spontaneously active interneurons and from the depression of the afterhyperpolarization [gK(Ca)]⁷. It is well known that reducing the concentration of divalent cations bathing a variety of nerve and muscle preparations increases excitability⁶⁻⁹, probably through their effect on surface charge density. Mg^{2+} is less effective than Ca^{2+} in this role, thus the 0.2 mM Ca^{2+} , 4 mM Mg^{2+} solution used in most of the present experiments

would be expected to have a net excitatory effect. The increased excitability may be an essential requirement for field bursts because they were abolished by raising Mg^{2+} to 6 mM, which counterbalances the excitatory effect of the low Ca^{2+} more closely⁶. However, we did not detect any change in the action potential threshold, which would be expected if the hyperexcitability were primarily due to deficiency of divalent cations. The conduction block and irreversible damage of neurones caused by Ca^{2+} -free solutions^{6,10} were not seen here and would not be expected with the moderate changes in divalent cations used.

The potential shift recorded during the field bursts consisted of a 10–15 mV depolarization of the pyramidal cell membrane; its field was most negative in the cell body layer, and was positive in the outer two-thirds of the apical dendrites. This is consistent with a depolarization due to K^+ , released by the excessive neuronal discharge, accumulating in the laminae around the pyramidal cell bodies. Using ion-sensitive electrodes, we have measured increases in K^+ activity in the pyramidal layer during, but not preceding, the bursts (three experiments). The peak K^+ activity reached 10 mM which approximates the ceiling of the physiological range¹⁵. If the

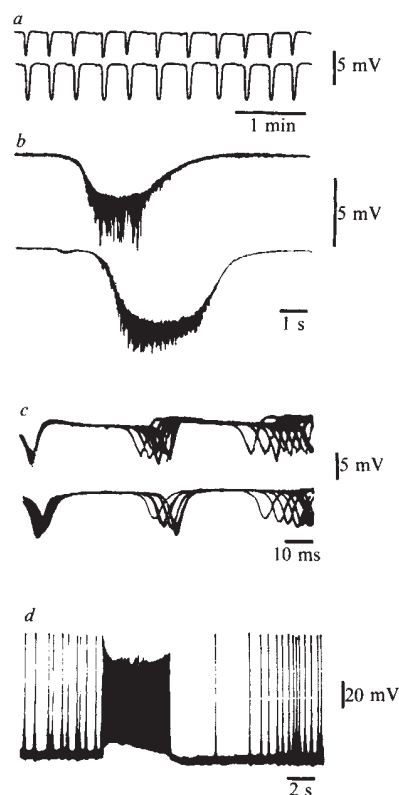


Fig. 1 Spontaneous 'field bursts' recorded from hippocampal slices which were incubated in 0.2 mM Ca^{2+} , 4 mM Mg^{2+} . The methods used have been described previously¹²⁻¹⁴. Briefly, 400 μm slices were prepared from rat hippocampus using a McIlwain tissue chopper or a Vibroslice and were incubated at 32 °C in a perfusion chamber in a thin film of perfusion liquid under warmed and moistened 95% O_2 /5% CO_2 . The standard media contained the following (in mM): Na^+ , 150; K^+ , 6.25; Cl^- , 137; Ca^{2+} , 2.0 or 0.2; Mg^{2+} , 2.0 or 4.0; HCO_3^- , 26; HPO_4^{2-} , 1.25; glucose, 10; and were equilibrated with 95% O_2 /5% CO_2 gas mixture to maintain the pH at 7.4. Micropipettes filled with 3 M NaCl (1–20 M Ω) or K-acetate (50–100 M Ω) were used for extra- and intracellular recording, respectively. *a*, Extracellular recordings from two sites, 1 mm apart, in the CA1 pyramidal layer. The negative field bursts recurred at 3 min⁻¹; *b*, their structure is shown in more detail on oscilloscope traces, population spikes descend from the negative potential shifts. *c*, Several superimposed sweeps during a field burst showing synchronization of population spikes. The recordings were made at sites 0.5 mm apart. *d*, Intracellular events corresponding to a field burst include a rapid, sustained burst of action potentials on a depolarizing shift, followed by a hyperpolarization.

resting potential were entirely due to K^+ this increase would correspond to a depolarization of 11 mV, which is close to the values we observed.

Treatments which depress the activity of ion pumps increased the frequency of the spontaneous bursts with little effect on their size or duration. Ouabain was effective at a concentration of 10^{-6} M; anoxia produced by reducing the gas flow over the slices also greatly speeded up the bursts (for example, from 0.2 min^{-1} to 10 min^{-1} within one minute of stopping the gas flow); cooling the slices from 32 to 25 °C significantly increased the burst frequency. A hyperpolarization was regularly recorded after the bursts, corresponding to a small, ouabain-sensitive, positive field in some extracellular recordings from the pyramidal layer, and may be attributed to the Na^+-K^+ pump. We postulate that sodium pump activity, stimulated by the increased extracellular K^+ (ref. 16), eventually terminates the burst by hyperpolarizing the cells for a period, and that the above treatments depress its ability to hyperpolarize without affecting its sensitivity to K^+ .

Several drugs and transmitter candidates, applied in nano- to micromolar concentrations, modified the interburst interval but not the burst size or duration. The frequency of field bursts was increased by the excitatory amino acids glutamate, aspartate and DL-homocysteic acid, and by acetylcholine, noradrenaline, histamine and Li^+ . The field bursts became less frequent during the application of the depressant amino acids γ -aminobutyric acid, taurine and imidazole acetic acid, and of serotonin, vasopressin and opiates. The stability of the phenomenon and its sensitivity to this wide range of drugs suggest that it may become a useful pharmacological tool (H.L.H. and J.G.R.J., in preparation).

The population spikes varied in amplitude during each field burst, producing characteristic shapes that varied between slices or even between different sites in CA1 of a single slice. However, our intracellular records showed a rapid and quite regular discharge. Thus for some periods during the field bursts, pyramidal neurones spread over substantial distances discharged in synchrony (Fig. 1c). The simplest hypothesis for the synchronization of the spikes is that extracellular currents, generated by pyramidal cells discharging, modulate the excitability of other members of the population, that is, 'field interaction'. Whether enough of the 'extracellular' loop of the action current passes through other cells to produce this effect depends on the detailed local geometry and electrical properties of the dendrites, glia and extracellular space. Experimental support for this possibility may be derived from the sensitivity of the smaller hippocampal granule cell to extrinsic electric fields¹³. The steady depolarization of the pyramidal cells in low Ca^{2+} would tend to increase the effectiveness of field interactions. The synchrony does not arise in the tightly-packed CA1 axons in the alveus as alvear fibre potentials were only seen after, and not preceding, the population spikes.

The synchronization of the bursts across the CA1 region remains to be explained. Paired recordings at sites 0.5–2.0 mm apart show that the leading site may vary between bursts. The latencies observed were up to 1 s, but more often ~50 ms. This is too fast to be explained solely by potassium movements of the kind involved in spreading depression, which propagates at a few mm per min¹⁷. We observed longer latencies, of the order of several seconds, between sites separated by an alvear lesion, so K^+ may transmit the field burst across these lesions. This also suggests that interactions within, or at least involving, the alveus contribute to the closer synchrony seen in the intact slice; but such interactions are not sufficient for synchronization, as lesions sparing the alveus alone result in two independent burst generators. Perhaps K^+ released by a bursting region can facilitate bursts in distant CA1 pyramidal cells more quickly by affecting their axons in the alveus than is possible through diffusion to cell bodies.

Alternatively, the burst synchronization, and even its generation, could be the result of specialized electrical contacts between the pyramidal neurones in a manner similar to that

proposed for recurrent excitatory chemical synapses in CA3 (ref. 1). Electrical synapses are sensitive to the removal of divalent cations¹⁰, although the small net changes involved in the present work might, if anything, be expected to weaken them. The existence of such contacts in CA1 is still controversial^{18,19} and even if they were involved here, we would still have to explain why the bursts do not occur in CA3 where similar connections have been described.

Further work is needed to determine whether the mechanisms responsible for the present phenomenon operate *in vivo*. While low Ca^{2+} may not be the only route to their expression, and one might speculate about anoxia or metabolic defects, it is of particular interest that similar conditions can prevail in the intact animal. Measurements with ion-selective electrodes have shown that Ca^{2+} levels fall before the start of seizures induced by pentylene tetrazol¹¹ and can drop to 0.5 mM. During these seizures K^+ increased to 7–8 mM, which would also have enhanced the mechanisms involved here.

Calvin has commented on the likelihood of there being a wide variety of bursting mechanisms available in the repertoire of neuronal properties²⁰. Previous studies of spontaneous synchronous bursting, notably in the hippocampal slice, have emphasized the role of synaptic connections. The phenomenon described here demonstrates how, in the absence of synaptic mechanisms, an increased neuronal discharge can be synchronized into a strongly rhythmic bursting behaviour.

J.G.R.J. thanks the Thorn Foundation for a Research Fellowship and the Wellcome Trust for a Travel Grant.

Received 25 June; accepted 8 October 1982.

1. Traub, R. D. & Wong, R. K. S. *Science* **216**, 745–747 (1982).
2. Schwartzkroin, P. A. & Prince, D. A. *Brain Res.* **147**, 117–130 (1978).
3. Wong, R. K. S. & Prince, D. A. *Science* **204**, 1228–1231 (1979).
4. Andersen, P. & Sears, T. A. *J. Physiol., Lond.* **173**, 459–480 (1964).
5. Calvin, W. H. *Brain Res.* **39**, 71–94 (1972).
6. Frankenhaeuser, B. & Hodgkin, A. L. *J. Physiol., Lond.* **137**, 218–244 (1957).
7. Hotson, J. R. & Prince, D. A. *J. Neurophysiol.* **43**, 409–419 (1980).
8. Orchardson, R. J. *J. Physiol., Lond.* **275**, 177–189 (1978).
9. Brismar, T. & Frankenhaeuser, B. *Acta physiol. scand.* **85**, 237–241 (1972).
10. Alvarez-Leefmans, F. J., de Santis, A. & Miledi, R. *J. Physiol., Lond.* **294**, 387–406 (1979).
11. Heinemann, U., Lux, H. D. & Gutnick, M. J. *Expl Brain Res.* **27**, 237–243 (1977).
12. Haas, H. L., Schaefer, B. & Vosmansky, M. *J. Neurosci. Meth.* **1**, 323–325 (1979).
13. Jefferys, J. G. R. *J. Physiol., Lond.* **319**, 143–152 (1981).
14. Jefferys, J. G. R. *J. Physiol., Lond.* **324**, 2P (1982).
15. Heinemann, U. & Lux, H. D. *Brain Res.* **120**, 231–249 (1977).
16. Baker, P. F. *et al. J. Physiol., Lond.* **200**, 459–496 (1969).
17. Leao, A. A. P. in *Experimental Models of Epilepsy* (eds Purpura, D. P., Penry, J. K., Tower, D., Woodbury, D. M. & Walter, R.) 173–196 (Raven, New York, 1972).
18. MacVicar, B. A., Ropert, N. & Krnjevic, K. *Brain Res.* **238**, 239–244 (1982).
19. Knowles, W. D. & Schwartzkroin, P. A. *J. Neurosci.* **1**, 318–322 (1981).
20. Calvin, W. H. in *Epilepsy: a Window to Brain Mechanisms* (eds Lockard, J. S. & Ward, A. A.) 97–121 (Raven, New York, 1980).

Intracellular recorded synaptic antagonism in the rat dentate gyrus

Vincenzo Crunelli, Susan Forda,
Graham L. Collingridge & John S. Kelly

Department of Pharmacology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

With the exception of acetylcholine at the motoneurone-Renshaw cell synapse, the excitatory chemical transmitters of the major pathways in the vertebrate brain are unknown¹. However, on the basis of biochemical and electrophysiological experiments at the extracellular level, the excitatory acidic amino acids glutamate and aspartate have long been suggested as the most likely candidates for this role^{2–4}. Unfortunately, difficulties with recording intracellularly from postsynaptic cells at many of these synapses and the lack of selective antagonists have curtailed the criteria by which the identification of glutamate and aspartate as transmitters can be explored. We have now made intracellular recordings from granule cells of the dentate gyrus *in vitro* and report that γ -D-glutamylglycine

(γ -DGG) is an effective antagonist of the excitatory postsynaptic potential (e.p.s.p.) evoked by stimulation of the medial perforant path (PP). During this antagonism the reversal level of the e.p.s.p. remains the same and the passive membrane properties of the postsynaptic membrane are unaltered. Quantal analysis shows that the reduction in the e.p.s.p. amplitude can be accounted for almost entirely by a reduction in quantal size while the quantal content remains unchanged, indicating a direct reduction in the postsynaptic sensitivity to the endogenous transmitter. ($\pm$ 2-Amino-5-phosphonovalerate (APV) proved to be ineffective as a blocker of the PP-evoked e.p.s.p. and *cis*-2,3-piperidinedicarboxylate (PDA) caused excitation. The selectivity of γ -DGG and APV against the depolarization induced by the iontophoretic application of *N*-methyl-D-aspartate (NMDA), quisqualate and kainate closely resembled that in spinal cord. This constitutes the first evidence from a readily identifiable synapse in support of the view that excitatory amino acid receptors of the quisqualate/kainate type are involved in synaptic transmission and offers further support for the suggestion that the endogenous transmitter is aspartate or glutamate.

It has been proposed that three types of pharmacologically distinct excitatory amino acid receptors can be distinguished in the spinal cord, classified according to their most potent agonists as NMDA, kainate and quisqualate⁵. Thus APV is the most specific and potent antagonist of NMDA receptors^{5,6}. γ -DGG, as well as blocking NMDA receptors, antagonizes responses to kainate^{5,7}, whereas PDA antagonizes all three receptor types^{5,8}. Using intracellular electrodes we have studied the effect of APV, γ -DGG and PDA on the synaptic excitation evoked by PP stimulation in the granule cells of the dentate gyrus⁹ where there is strong evidence favouring glutamate as the endogenous transmitter¹⁰⁻¹³.

Impaled neurones, in a standard hippocampal slice preparation (see Fig. 1 legend), were characterized as granule cells by their position in the granular layer of the dentate gyrus, by orthodromic activation from the PP and by antidromic invasion from the mossy fibres region of CA₃^{14,15}. Experiments were carried out on only those granule cells whose resting membrane potential (< -60 mV) and membrane input resistance remained stable throughout the experiment. Stimulation of the PP evoked an e.p.s.p. at least 8–20 mV in amplitude yet still below the threshold for initiating action potentials. This may be explained in part by the marked decrease in membrane resistance ($46 \pm 3\%$, $n = 6$) shown to be associated with the peak of the e.p.s.p. by computing voltage-current plots off-line with a PDP 11/23 computer (Cambridge Electronic Design Ltd). The same computer program was used to compute the apparent reversal level of the e.p.s.p. (-14.5 ± 3.4 mV, $n = 4$) by extrapolation from voltage-current plots determined from changes in peak e.p.s.p. amplitude (located mathematically) produced by the intracellular injection of long pulses (200 ms) of depolarizing and hyperpolarizing current. The voltage-current plots showed no significant deviation from linearity indicative of attenuation in e.p.s.p. amplitude caused by significant changes in the time constant of the cell soma or principal dendrites, or by the intracellular injection of current or extracellular application of amino acids.

On 29 granule cells, the iontophoretic application of γ -DGG (25–275 nA) markedly and reversibly reduced the amplitude of the e.p.s.p. evoked by stimulation of the PP (Fig. 1). The mean reduction of the e.p.s.p. was 61% (range 38–80%) of control. The effect of γ -DGG was not associated with a change in the resting membrane potential. Computer-derived voltage-current plots showed the resting membrane input resistance to be unaltered as was the excitability of the cell, tested both by determining the voltage threshold required for antidromic invasion and by the direct injection of depolarizing current.

These results do not exclude the possibility that the antagonism could be due to a presynaptic action of γ -DGG. However, in 12 experiments in which the mass presynaptic fibre volley could be recorded by our intracellular electrode (Fig. 1), its

amplitude and duration were unaltered by the application of γ -DGG, even at the highest iontophoretic currents used (275 nA).

Independent evidence in support of the view that the action of γ -DGG on the postsynaptic membrane is competitive, came from a more detailed analysis of the e.p.s.p. For example, in five of six cells, the voltage-current plots showed the apparent reversal potential of the e.p.s.p. to be unaltered by the iontophoretic application of γ -DGG, that is, no systematic alteration occurred in the point of intersection of the extrapolated voltage-current plots derived on and off the peak of the e.p.s.p. in the presence and absence of γ -DGG. In effect, the decline in amplitude of the peak e.p.s.p. was accompanied by an increase in membrane resistance. This continuous linear relationship between the amplitude of the e.p.s.p. and membrane resistance is one prediction of the current theory that the amplitude of the e.p.s.p. is determined by the summation of individual currents generated by the opening of a relatively constant number

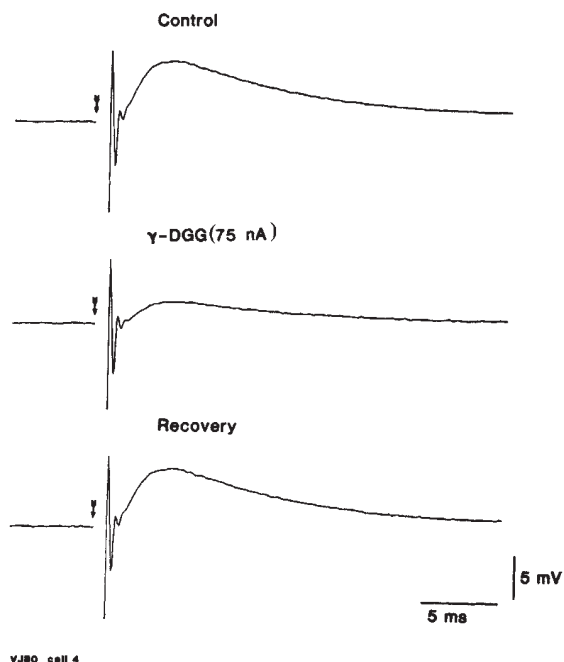


Fig. 1 Reversible reduction by γ -D-glutamylglycine (γ -DGG) of the medial perforant path-evoked e.p.s.p. recorded in a granule cell of the dentate gyrus *in vitro*. After impalement, the position of the iontophoretic electrode was adjusted along the dendritic tree at the level of the middle molecular layer until the response to a short (~ 1 s) application of glutamate was maximal. Glass micropipettes (tip diameter 3–6 μ m) filled with 1 M NaCl were used as stimulating electrodes and positioned in the medial perforant path, 4–6 mm away from the impaled granule cell. Each trace is the computer average of 60 successive e.p.s.p.s recorded before (control), during (γ -DGG) and after (recovery) the iontophoretic application of 75 nA of γ -DGG. The e.p.s.p. was reversibly reduced from 8.5 to 3.2 mV. Our intracellular electrode recorded the presynaptic fibre volley which was unaltered by the application of γ -DGG. Arrows mark the preceeding stimulus artefact. Hippocampal slices (400- μ m thick) were prepared from decapitated rats (200 g) using a McIlwain tissue chopper and maintained *in vitro* in a warm (36 $^{\circ}$ C) continuously oxygenated (95% O₂, 5% CO₂) medium containing (mM): NaCl 134, KCl 5, KH₂PO₄ 1.25, MgSO₄ 2, CaCl₂ 2, NaHCO₃ 16 and glucose 10. Intracellular microelectrodes were filled with 1 M potassium acetate and had a resistance of 80–120 M Ω when measured *in situ* with a WPI model 707 electrometer. Drugs were applied by passing negative current through a multibarrelled micropipette containing L-glutamate (1 M, pH8), γ -DGG (0.2 M, pH8), APV (0.05 M in 100 mM NaCl pH8), PDA (0.2 M, pH8), NMDA (0.02 M in 150 mM NaCl, pH8), quisqualate (0.02 M in 150 mM NaCl, pH8), kainate (0.02 M in 150 mM NaCl, pH8) and NaCl (1 M, pH8). A positive backing current of 1–20 nA was applied when necessary.

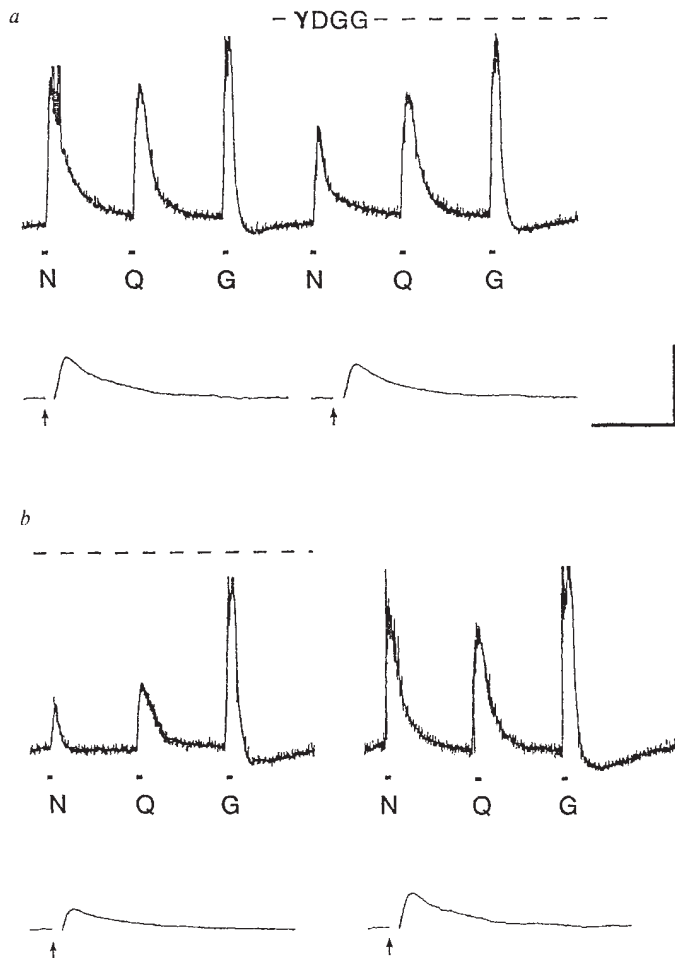


Fig. 2 Intracellular voltage recordings from a granule cell showing the effect of γ -DGG on the medial PP-evoked e.p.s.p. and the depolarization induced by pulsed iontophoretic applications of NMDA (N), quisqualate (Q) and glutamate (G). The upper trace is the intracellular voltage response to drug application, and the lower trace the computer average of 10 e.p.s.p.s. In *a* a 13-s application of γ -DGG was sufficient to reduce the NMDA-induced depolarization to 59% of control, while the e.p.s.p. was unaltered. In *b* the continuing application of γ -DGG reduced the response to quisqualate and the e.p.s.p. amplitude decreased from 11 to 5 mV. *a* and *b* are consecutive records and the gap in *b* is 1.5 min. The action potentials evoked by the drug-induced depolarization have been attenuated by the computer plotting subroutines. The durations of the amino acid applications are indicated by the bars below the voltage records. Drug ejection currents were (nA): NMDA 100, quisqualate 150, glutamate 200 and γ -DGG 150. Calibration bars represent 20 mV and 30 s for the upper trace, 20 mV and 32 ms for the lower one. Solid arrows mark stimulation of the PP.

of transmitter-specific channels¹⁶, and that antagonists simply combine with specific receptors to reduce the number of channels available. As reported below (Fig. 2) we also tested a further prediction of this theory by exploring the possibility that the potential evoked by the exogenous application of the transmitter involves similar if not identical receptor-activated channels.

In four cells in which all the criteria necessary to carry out a quantal analysis by the variance method^{17,18} on trains of e.p.s.p.s recorded in the presence and absence of γ -DGG were met, the decrease (46%) in the e.p.s.p. amplitude could be accounted for almost entirely by a reduction in quantal size (control, 294 μ V; γ -DGG, 133 μ V; $n=4$) while the quantal content remained unchanged (control, 46; γ -DGG, 50;

$n=4$)¹⁹. Thus, according to the quantal model of synaptic transmission²⁰, the effect of γ -DGG on the PP-evoked e.p.s.p. is explained by a decreased sensitivity of the postsynaptic membrane to the endogenous transmitter without any appreciable change in transmitter release.

APV, the most potent and selective NMDA-receptor antagonist in the spinal cord^{5,6}, was found to be inactive as an antagonist of the PP-evoked e.p.s.p. In eight of nine neurones on which γ -DGG reversibly reduced the e.p.s.p., the iontophoretic application of APV (50–300 nA) was ineffective. In the only granule cell in which it appeared to be active, the decrease in the e.p.s.p. amplitude was not associated with any change in membrane potential or resting input resistance.

PDA, tested with iontophoretic current ranging from 1 to 50 nA, evoked an excitation of granule cells (six of eight neurones) accompanied by a depolarization and an increase in membrane resistance. Even at low iontophoretic currents, PDA produced burst-like activity.

The receptor and channels antagonized by γ -DGG, were characterized by comparing the action of γ -DGG on the PP-evoked e.p.s.p. with the amino acid-induced depolarization of granule cells. As shown in Fig. 2, the reduction in the e.p.s.p. amplitude only occurred when the application of γ -DGG was of sufficient magnitude to reduce the depolarization evoked by quisqualate and/or kainate, even though much smaller applications could block the response produced by NMDA. In the same experiments, we confirmed the even greater selectivity of APV for the NMDA receptor and its inability to effectively block the response induced by quisqualate and kainate.

In another series of experiments, short applications of γ -DGG but not APV were shown to antagonize the PP-evoked e.p.s.p. and glutamate-induced depolarization. However, when the action of γ -DGG on glutamate- as opposed to aspartate-induced depolarization was examined, no clear pattern of selectivity was observed. This was at variance with previous results in the spinal cord^{5,7}, in fact, in all six granule cells tested, γ -DGG was unselective and reduced the action of both excitants.

The results presented here show that γ -DGG is a potent antagonist of the PP-evoked e.p.s.p. and acts directly on the postsynaptic membrane to decrease its sensitivity to the endogenous transmitter. The amino acid receptors involved at this synapse are of the quisqualate and/or kainate type.

The work was supported by MRC award G8011874. We thank Dr J. C. Watkins (Bristol University) for the supply of antagonists, the Wellcome Trust for financial support to V.C. and G.L.C., and the SRC and ICI for a CASE award to S.F. We also thank Mr J. Wise of the Computer Network Service for his help and enthusiasm and Mrs A. Keable for typing the manuscript.

Received 20 August; accepted 6 October 1982.

1. Eccles, J. C. *The Physiology of Synapses* (Springer, Berlin, 1964).
2. Johnson, J. L. *Prog. Neurobiol.* **10**, 155–202 (1978).
3. Curtis, D. R. in *Glutamic Acid: Advances in Biochemistry and Physiology* (eds Filer, L. J., Garattini, S., Kare, M. R. & Reynolds, W. A.) 163–175 (Raven, New York, 1979).
4. Krnjević, K. & Phillis, J. W. *J. Physiol., Lond.* **165**, 274–304 (1963).
5. Watkins, J. C. & Evans, R. H. A. *Rev. Pharmac. Tox.* **21**, 165–204 (1981).
6. Davies, J. & Watkins, J. C. *Brain Res.* **235**, 378–386 (1982).
7. Davies, J. & Watkins, J. C. *Brain Res.* **206**, 172–177 (1981).
8. Davies, J., Evans, R. H., Francis, A. A., Jones, A. W. & Watkins, J. C. *J. Neurochem.* **36**, 1305–1307 (1981).
9. McNaughton, B. L. & Barnes, C. A. *J. comp. Neurol.* **175**, 439–454 (1977).
10. Spencer, H. J., Gribkoff, V. K., Cotman, C. W. & Lynch, G. *Brain Res.* **105**, 471–481 (1976).
11. Wheal, H. W. & Miller, J. J. *Brain Res.* **182**, 145–155 (1980).
12. Nadler, J. V., White, W. F., Vaca, K. W., Redburn, D. A. & Cotman, C. W. *J. Neurochem.* **29**, 279–290 (1977).
13. Sandoval, M. E., Horch, P. & Cotman, C. W. *Brain Res.* **142**, 285–299 (1978).
14. Andersen, P., Holmqvist, B. & Voorhoeve, P. E. *Acta physiol. scand.* **66**, 448–460 (1966).
15. Lomo, T. in *Excitatory Synaptic Mechanisms* (eds Andersen, P. & Jansen, J. K. S.) 207–211 (Universitetsforlaget, Oslo, 1970).
16. Ginsborg, B. L. *Pharmac. Rev.* **19**, 289–316 (1967).
17. Hubbard, J. I., Llinas, R. & Quastel, D. M. S. *Electrophysiological Analysis of Synaptic Transmission* (Arnold, London, 1969).
18. Kuno, M. *Physiol. Rev.* **51**, 647–678 (1971).
19. McNaughton, B. L., Barnes, C. A. & Andersen, P. *J. Neurophysiology*, **46**, 952–966 (1981).
20. Katz, B. *Nerve, Muscle and Synapse* (McGraw Hill, New York, 1966).

to reduce the map complexity but increase the specificity, because of the lower proportion of tyrosine compared with lysine residues in VSGs¹. The ¹²⁵I-maps (Fig. 2*b, c*) again demonstrated similarity between WaTat 1.1 and 1.12 VSGs, but at least four out of the total 15–19 labelled tryptic peptides were dissimilar. As a control the peptide maps of WaTat 1.1 and 1.12 VSGs were compared to that of WaTat 1.2, a VSG from a third, randomly derived trypanosome clone but no similarities were observed (Fig. 2*b, c*). Thus these results agreed with the isoelectrofocusing data showing that WaTat 1.1 and WaTat 1.12 VSGs were not identical. The two VSGs did, however, share amino acid sequences which were not present in a third VSG.

Seven monoclonal antibodies were prepared against WaTat 1.1 VSG either by immunization of mice with isolated VSG or by immunization with irradiated (40,000R) trypanosomes. All seven antibodies bound ¹²⁵I-labelled WaTat 1.1 VSG but only five reacted in immunofluorescence with live WaTat 1.1 trypanosomes (Table 1). Therefore, not all the epitopes recognized in the isolated VSG were equally accessible in the same VSG on the surface of live trypanosomes. Of the five monoclonal antibodies which were reactive against WaTat 1.1 trypanosomes in immunofluorescence, four also reacted with WaTat 1.12 trypanosomes. One monoclonal antibody (Tryp 22A₁) reacted only with WaTat 1.1 trypanosomes, showing that

an epitope exposed to antibody on WaTat 1.1 trypanosomes was not present on WaTat 1.12 trypanosomes. No evidence for population heterogeneity was obtained with these monoclonal antibodies, that is, when immunofluorescence reaction occurred, greater than 99% of the trypanosomes in that population reacted. Similarly, monospecific rabbit antisera against WaTat 1.12 VSG reacted with greater than 99% of both WaTat 1.1 and 1.12 trypanosomes.

As some, but not all, exposed epitopes were shared between WaTat 1.1 and 1.12 trypanosomes, we investigated whether cross-neutralization would occur. Results in Table 1 demonstrate that rabbit antiserum to either WaTat 1.1 or WaTat 1.12 VSGs neutralized both WaTat 1.1 and 1.12 trypanosomes and had no effect on WaTat 1.2 organisms. The two control sera (anti-WaTat 1.2 VSG and normal rabbit serum) were without effect on WaTat 1.1 or 1.12 trypanosomes.

We next analysed immunological cross-reactions between isolated WaTat 1.1 and 1.12 VSGs. Previous data have shown that homologous reactions between VSGs and monospecific rabbit anti-VSG sera were not effectively inhibited by heterologous VSGs⁴ except in the case of 'isotypic' trypanosomes⁵. In radioimmunoassay, the homologous reactions ¹²⁵I-WaTat 1.1 VSG versus rabbit anti-WaTat 1.1 VSG serum or ¹²⁵I-WaTat 1.12 VSG versus rabbit anti-WaTat 1.12 VSG serum were effectively inhibited by both WaTat 1.1 and 1.12

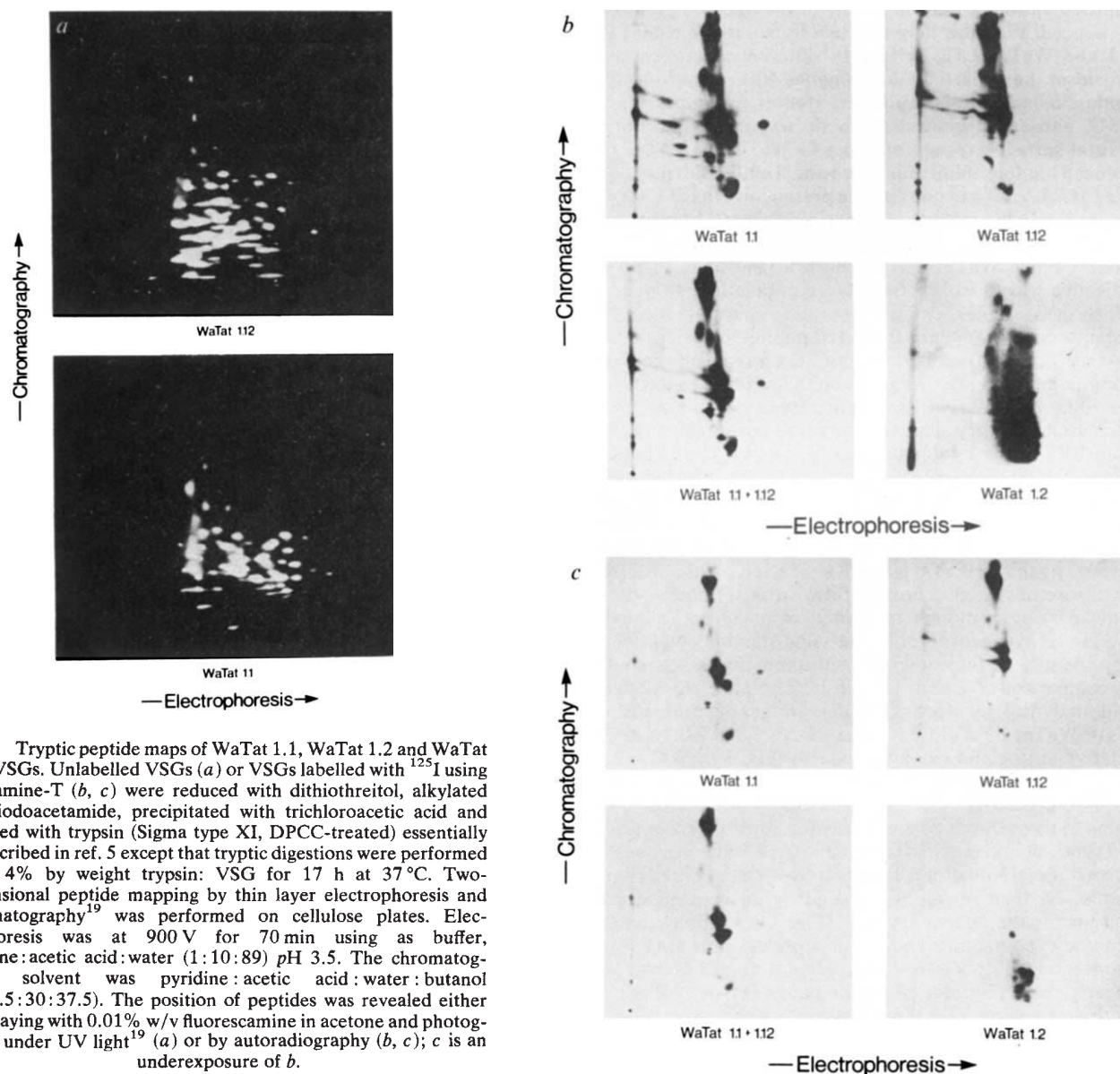


Fig. 2 Tryptic peptide maps of WaTat 1.1, WaTat 1.2 and WaTat 1.12 VSGs. Unlabelled VSGs (*a*) or VSGs labelled with ¹²⁵I using chloramine-T (*b, c*) were reduced with dithiothreitol, alkylated with iodoacetamide, precipitated with trichloroacetic acid and digested with trypsin (Sigma type XI, DPCC-treated) essentially as described in ref. 5 except that tryptic digestions were performed using 4% by weight trypsin: VSG for 17 h at 37°C. Two-dimensional peptide mapping by thin layer electrophoresis and chromatography¹⁹ was performed on cellulose plates. Electrophoresis was at 900 V for 70 min using as buffer, pyridine:acetic acid:water (1:10:89) pH 3.5. The chromatography solvent was pyridine:acetic acid:water:butanol (25:7.5:30:37.5). The position of peptides was revealed either by spraying with 0.01% w/v fluorescamine in acetone and photography under UV light¹⁹ (*a*) or by autoradiography (*b, c*); *c* is an underexposure of *b*.

Table 1 Reactivity of antibody with different trypanosomes and VSGs

Antibody	WaTat 1.1			WaTat 1.12			WaTat 1.2		
	RIA	FA	N	RIA	FA	N	RIA	FA	N
Tryp 1E ₁ , monoclonal	+	+	ND	+	+	ND	—	—	ND
Tryp 2B ₁ , monoclonal	+	+	ND	+	+	ND	—	—	ND
Tryp 4C ₁ , monoclonal	+	+	ND	+	+	ND	—	—	ND
Tryp 22A ₁ , monoclonal	+	+	ND	—	—	ND	—	—	ND
WaT 8A ₁ , monoclonal	+	—	ND	+	—	ND	—	—	ND
WaT 20A ₁ , monoclonal	+	—	ND	+	—	ND	—	—	ND
WaT 24A ₁ , monoclonal	+	+	ND	+	+	ND	—	—	ND
Rabbit anti-WaTat 1.1 VSG	+	+	0/10	+	+	1/10	+	—	10/10
Rabbit anti-WaTat 1.12 VSG	+	+	0/10	+	+	0/10	+	—	10/10
Rabbit anti-WaTat 1.2 VSG	+	—	9/9	+	—	8/10	+	+	0/9
Normal rabbit serum	—	—	10/10	—	—	9/10	—	—	10/10

Monospecific rabbit anti-VSG sera were prepared as in ref. 4. Mouse monoclonal antibodies were prepared as described^{20,21} using the SP2/0-Ag 14 myeloma cell line. The spleen cells for fusion were from mice immunized either with irradiated (40,000 R) WaTat 1.1 trypanosomes (Tryp series) or purified WaTat 1.1 VSG (WaT series). All hybridomas were double cloned. Radioimmunoassay (RIA) was as in Fig. 3 measuring binding of purified, ¹²⁵I-labelled VSGs to antibody. Immunofluorescence (FA) procedures were similar to those of ref. 4; the fluorescence of living trypanosomes was recorded immediately after the last wash. Neutralization (N) experiments were performed as follows. Ten trypanosomes in 50 µl of PSG (57mM Na₂HPO₄, 3mM NaH₂PO₄, 45mM NaCl, 55mM glucose, pH 8.0) were incubated with 100 µl of rabbit antiserum (inactivated at 56 °C for 30 min) for 30 min on ice and then injected into the peritoneal cavity of a mouse. Parasitaemia was observed by sampling tail blood at 2-day intervals. Numbers represent number of mice dead/total injected; all deaths were the result of trypanosome infection.

unlabelled VSGs. Inhibition of 100%, however, was only achieved by the homologous antigen. The heterologous antigen inhibited by a maximum of 70–90% (data not shown). A reaction of partial identity was also obtained between WaTat 1.1 and 1.12 VSGs in immunodiffusion.

The seven monoclonal antibodies made to WaTat 1.1 VSG were tested for binding to ¹²⁵I-labelled VSGs (Table 1 and Fig. 3). Six out of the seven bound both WaTat 1.1 and 1.12 VSGs. Monoclonal antibody Tryp 22A₁ only reacted with WaTat 1.1 VSG, a similar result to that obtained with the same antibody and live trypanosomes in immunofluorescence. Even though the remaining six monoclonal antibodies bound both WaTat 1.1 and 1.12 labelled VSGs, with four out of the six, more antibody was required to bind a given quantity of WaTat 1.12 VSG than 1.1 VSG. This indicated reduced binding constants of these four monoclonal antibodies for WaTat 1.12 VSG. The binding constant of antibody Tryp 4C₁ for WaTat 1.12 VSG was so low that binding was barely observable, even with undiluted antibody. None of the monoclonal antibodies bound ¹²⁵I-WaTat 1.2 VSG.

The data show that WaTat 1.1 and WaTat 1.12 trypanosomes: (1) are derived from a single trypanosome in a time period comparable with that available for variation in one host animal under field conditions; (2) express VSGs which are not identical in amino acid sequence, but do share regions of amino acid sequence homology not found in other, randomly isolated VSGs; (3) contain both similar and different surface-exposed VSG epitopes. We consider that the most likely mechanisms by which WaTat 1.1 trypanosomes differentiated into WaTat 1.12 are the following: (i) expression of alternative genes (perhaps members of the same VSG gene family^{6,7}) present in WaTat 1.1 organisms; (ii) re-expression of the same WaTat 1.1 VSG gene which has undergone rearrangement⁸ or mutation^{9,10} either before or during re-expression; (iii) a combination of both mechanisms (i) and (ii), that is, expression of alternative genes and rearrangement or mutation. It may be possible to discriminate between these possibilities by analysing for the presence of WaTat 1.1 and 1.12 genes in both trypanosomes.

It might be argued that oligosaccharides could be responsible for the cross-reactions observed between WaTat 1.1 and WaTat 1.12 VSGs. We do not consider this likely because: (1) oxidation of WaTat 1.1 VSG with 100 mol periodate per mol VSG (48 h, 4 °C, as in ref. 11) destroyed binding of the antigen to rabbit anti-WaTat 1.2 VSG sera (due to cross-reacting oligosaccharide^{11,12}), but did not affect binding of the oxidized WaTat 1.1 VSG to rabbit anti-WaTat 1.12 VSG sera; (2) various studies using periodic acid-Schiff staining and electron microscopy, immunochemical techniques and sequencing^{11–17} suggest that VSG carbohydrate is on the innermost side of the

surface coat next to the cytoplasmic membrane, and is not accessible for binding to lectins or antibody when on the trypanosome surface. The monoclonal antibodies shown in Fig. 3 react with surface-accessible epitopes.

These results have theoretical and practical implications. First, a single trypanosome may vary during a short time period into a population expressing a different VSG, from which it might not be distinguished using monospecific or even some monoclonal anti-VSG sera in immunofluorescence. We suggest that amino acid sequence identity between expressed VSGs be the ultimate criterion of whether two trypanosome populations are identified as being similar. The data also provide the first steps to a classification of different trypanosome populations depending on shared, exposed epitopes present on their surface. This classification would be useful in deciding whether cross-immunization between any two trypanosome populations was possible.

Second, reappearance of the same antigenic type of trypanosome at different times during a cattle infection has been suggested previously¹⁸. This phenomenon could, clearly, now also be attributed to appearance of a trypanosome expressing a similar but not identical VSG.

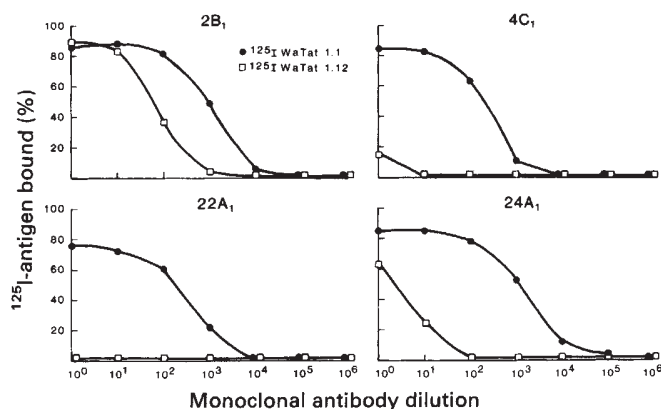


Fig. 3 Binding of monoclonal antibodies to ¹²⁵I-labelled WaTat 1.1 and WaTat 1.12 VSGs. Reaction mixtures contained 20 µl of ¹²⁵I-labelled VSGs⁴ (5ng in all cases), 20 µl of monoclonal antibody and 20 µl of TEN buffer (0.02 M Tris-HCl, 0.005 M EDTA, 0.1 M NaCl, 0.015 M NaN₃, pH 7.6 and 5 mg ml⁻¹ bovine serum albumin). The mixture was incubated for 1 h at room temperature then 100 µl of a 10% suspension of formalin-fixed *Staphylococcus aureus* (Pansorbin, Calbiochem) was added. The incubation continued for another hour before the mixture was centrifuged at 1,500g for 20 min, washed twice in TEN buffer and radioactivity in the final pellet measured. All dilutions were made in TEN buffer.

Finally, these two trypanosome populations represent the first available for analysing rearrangements and mutations in VSG genes and relating DNA sequence changes to modifications in individual, surface-exposed epitopes.

We thank Eldon Libstaff, Jeff Regnier, Richard Dixon and Sandra Seifert for technical assistance. Research support was provided by the state of Washington Initiative Measure no. 171 and by Rockefeller Foundation grant GA-COH-8013.

Received 5 July; accepted 30 September 1982.

1. Cross, G. A. M. *Parasitology* **71**, 393–417 (1975).
2. Bridgen, P. J., Cross, G. A. M. & Bridgen, J. *Nature* **263**, 613–614 (1976).
3. Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **294**, 53–57 (1981).
4. Barbet, A. F. & McGuire, T. C. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1989–1993 (1978).
5. Vervoort, T. *et al. Immunology* **44**, 223–232 (1981).
6. Hoeijmakers, J. H. J. *et al. Nature* **284**, 78–80 (1980).
7. Pays, E., Van Meirvenne, N., LeRay, D. & Steinert, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2673–2677 (1981).
8. Michels, P. A. M., Bernards, A., Van der Ploeg, L. H. T. & Borst, P. *Nucleic Acids Res.* **10**, 2353–2366 (1982).
9. Frasch, A. C. C., Borst, P. & Van den Burg, J. *Gene* **17**, 197–211 (1982).
10. Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **298**, 676–679 (1982).
11. Barbet, A. F., Musoke, A. J., Shapiro, S. Z., Mpimbaza, G. & McGuire, T. C. *Parasitology* **83**, 623–637 (1981).
12. Holder, A. A. & Cross, G. A. M. *Molec. Biochem. Parasit.* **2**, 135–150 (1981).
13. Wright, K. A. & Hales, H. J. *Parasitology* **56**, 671–683 (1970).
14. Cross, G. A. M. *Nature* **277**, 310–312 (1979).
15. Johnson, J. G. & Cross, G. A. M. *Biochem. J.* **178**, 689–697 (1979).
16. Seed, T. M., Seed, J. R. & Brindley, I. *Tropenmedizin und Parasitenkunde* **27**, 202–212 (1976).
17. Steiger, R. F. *Acta trop.* **32**, 152–158 (1975).
18. Nantulya, V. M., Musoke, A. J., Barbet, A. F. & Roelants, G. E. J. *Parasit.* **65**, 673–679 (1979).
19. Whittaker, R. G. & Moss, B. A. *Analyt. Biochem.* **110**, 56–60 (1981).
20. Davis, W. C. *et al. Proc. 7th natn. Anaplasmosis Conf.*, 1981 (in the press).
21. Shulman, M., Wilde, C. O. & Köhler, G. *Nature* **276**, 269–270 (1978).

A human suppressor T cell clone which recognizes an autologous helper T cell clone

Jonathan R. Lamb* & Marc Feldmann†

Division of Immunologic Oncology, Lombardi Cancer Research Center, Georgetown University Medical School, Washington DC 20007, USA

† Imperial Cancer Research Fund Tumour Immunology Unit, Department of Zoology, University College London, Gower Street, London WC1E 6BT, UK

In contrast to the many reports of IL-2 dependent proliferating, helper or killer cell clones¹, there is only a single report of an IL-2 dependent suppressor cell clone², in the mouse. However by the same cloning procedures used to generate human helper cells³, suppressor cell clones to influenza virus could not be generated, and so another strategy was used. Jerne's network hypothesis⁴ proposes that immune regulation results from lymphoid cell receptors recognizing determinants on other lymphoid cell receptors. If this is the case it should be possible to generate regulatory T cell clones against other T cells and we report here the generation of an autologous suppressor cell clone which recognizes and inhibits the function of a human helper T cell clone. Such an autologous suppressor cell clone provides a new approach to understanding the pathways and molecular events involved in immune regulation. Mutual stimulation of the suppressor cell clone and the target helper cell clone in the absence of back stimulation, provides direct experimental evidence for the existence of interactions between T cell receptors, and thus suggests that the specificity of the suppressor cell clone is for the antigen receptor of the helper cell.

Human immunology, unlike mouse immunology, is limited by restricted access to genetically uniform cells. In order to circumvent this problem, all the experiments reported here were performed with cryopreserved cells from a single

individual. T lymphocyte clone 6 (TLC 6) is a helper cell which recognized influenza A matrix protein in conjunction with HLA-DR1^{3–6}. It was used to generate anti-idiotypic suppressor cells because it grows well in culture, and its helper activity is easy to measure *in vitro* using the influenza virus specific antibody response and assay system developed by our colleagues, Callard, Beverley and Zanders^{7–9}, using either the cells themselves, or its antigen specific helper factor.

To stimulate autologous peripheral blood cells (PBL) with irradiated TLC6, a 10-day mixed lymphocyte culture was performed. The resulting blasts were purified on Ficoll-Hypaque and restimulated with irradiated TLC6. The ensuing lymphoblasts purified on a Percoll gradient three days later, and cloned as described in Table 1. Approximately 60 clones were grown up, and tested for their specificity. For simplicity, proliferation was used as the criterion of responsiveness, and cells from individual clones were cultured with irradiated autologous PBL and either the immunogen TLC6, another helper clone from the same donor HA1.9, influenza A virus, or just PBL alone. After 72 hours, cultures were pulsed with tritiated thymidine and cells were collected 16 hours later. Several patterns of reactivity were seen. Clones from these experiments were termed AC1 to 60. The most common response was to influenza virus (for example clone AC10), which could have been still present on the surface of washed TLC6 even after one week's culture. Other clones showed no detectable response to any of the stimuli used (for example AC57). This is paradoxical, as only activated cells respond to TCGF and can be cloned¹. It seems likely that these clones were responding to TLC6, but the response was not detectable by the assay used. Clone AC34 is presumably recognizing an antigen on PBL (or serum), whereas clone AC5 is recognizing an activated cell antigen. The specificity of clones AC19 and AC22 is not known.

The most relevant clone, chosen for further analysis, was clone AC50. This proliferated in response to TLC6, but not other clones, PBL or influenza virus. Thus it seemed to be specific for TLC6, and was a candidate for an anti TLC6 clone. In the absence of any antibodies which unambiguously define human T-cell receptors, we used a functional assay to analyse the specificity of clone AC50 further.

We argued that if receptors on AC50 recognized anti-TLC6 receptors, then TLC6 might be stimulated to respond to irradiated AC50 *in vitro*, as its receptors would be bound. Thus a variety of influenza specific clones (all from the same donor) were stimulated with irradiated clone AC50, or AC50 plus irradiated E⁺ cells (T depleted cells) as a source of antigen presenting cells. As a control the response to AC50, PBL and influenza virus was assayed. As shown in Table 2, all the clones responded to influenza virus and TCGF, but only TLC6 responded to AC50, with or without the addition of irradiated E⁺ cells. This result is compatible with the notion that AC50 is an auto anti-idiotypic clone, that is, recognizes antigen-specific receptors on TLC6. The response in the absence of additional antigen presenting cells (APC) is of interest, implying that the reaction between TLC6 and AC50 may not require antigen presenting cells. However, a few APC, which had not adhered to the plastic in the week of culture since irradiated PBL were added could still have been present, and this question needs to be analysed in detail. It is of interest that Infante *et al.*¹⁰ recently found that anti-idiotypic antisera stimulate mouse T cell clones to proliferate even in the absence of added antigen presenting cells, in analogy to the results shown here with anti-idiotypic T cells.

The results in Table 2 indicate that AC50 and TLC6 interact in a specific manner. But they do not conclusively indicate that the reaction is anti-receptor. An alternative interpretation is that the proliferation of TLC6 is due to back stimulation^{3,11} with irradiated AC50 recognizing TLC6, and it is the release of lymphokines from irradiated AC50 which causes TLC6 to incorporate thymidine. This possibility was investigated, using a cultured (not cloned) T lymphocyte line (CTLL) which responds well to TCGF (Table 3). A mixture of irradiated AC50

* Present address: Imperial Cancer Research Fund, Human Tumour Immunology Unit, University College Hospital, London WC1E 6BT, UK.

Table 1 Induction of anti-TLC 6 clones

Clone no.	Antigen					
	6 + PBL	HA1.9 + PBL	PBL	Flu + PBL	Medium	TCGF
AC5	1,249 ± 94	5,469 ± 1,595	44 ± 10	75 ± 21	39 ± 16	11,450 ± 750
AC10	43 ± 20	57 ± 16	54 ± 16	4,676 ± 404	31 ± 11	5,063 ± 162
AC19	2,534 ± 367	2,794 ± 366	2,457 ± 326	31 ± 8	15 ± 4	3,164 ± 243
AC22	1,518 ± 247	24 ± 2	46 ± 11	2,478 ± 35	34 ± 4	8,695 ± 622
AC34	1,801 ± 360	2,255 ± 304	1,655 ± 68	3,005 ± 912	29 ± 7	6,243 ± 221
AC50	2,928 ± 94	32 ± 13	36 ± 10	54 ± 8	21 ± 3	9,949 ± 216
AC57	23 ± 4	22 ± 5	24 ± 4	30 ± 5	21 ± 2	9,720 ± 506
AC58	4,550 ± 570	3,711 ± 446	7,508 ± 1,620	5,801 ± 1,494	26 ± 3	9,097 ± 208

Peripheral blood lymphocytes (PBL, 5×10^5 per ml) were cultured with irradiated (2,500 rad; ^{137}Cs) TLC6 (5×10^5 per ml) in RPMI 1640 supplemented with 10% pooled A⁺ serum in a 24 well plate (Costar). Both the responder (PBL) and stimulator (TLC6) cells were derived from the same donor. After 10 days of primary culture the viable cells were isolated on Ficoll-Hypaque and blast cells (10^5 per ml) were restimulated with irradiated TLC6 (10^5 per ml) in the presence of autologous irradiated PBL (5×10^5 per ml). Three days after restimulation the lymphoblasts were enriched on a discontinuous Percoll (Pharmacia) density gradient (35–40%) and plated at 0.3 cells per well in sterile 60 well Microtest II trays (Falcon) in the presence of T cells growth factor (TCGF or Interleukin 2) irradiated stimulator cells (TLC6; 5×10^5 per well) and autologous irradiated PBLs (5×10^5 per cell) as a source of antigen presenting cells (APCs). TCGF was prepared from PBL (1×10^6 per ml) cultured with PHA (0.1%) in the presence of 1% autologous serum for 48 hours as described elsewhere^{3,11}. The wells that showed positive growth (3.4% of the total number of blast cell seeded) at 7 days were transferred to 96 well flat bottom microtitre trays (Falcon) and cultured with irradiated TLC6 (5×10^4 per well) and autologous irradiated PBL (5×10^4 per well) in the presence of TCGF. Following a further 7 days in culture, the clones were transferred to 24 well trays containing TCGF, irradiated TLC6 (5×10^4 per well) and autologous irradiated PBL (1×10^6 per well). In the preliminary screen clones were examined for antigen specificity in a 72 hour array by using titrated methyl thymidine (^3H -TdR; NEN). A 1:200 dilution of individual clones from 24 well trays were cultured with autologous irradiated PBL (25×10^3 per well) and irradiated TLC6 (5×10^3 per well) in a total volume of 200 μl . Clones T cells from the same donor but with a different antigen specificity (HA1.9) together with autologous PBL, PBL alone or influenza A virus (A Texas/1/77; 5 haemagglutinating units (HAU) per ml) in the presence of PBL were added as controls. After 72 hours, cultures were pulsed with 1.0 μCi of ^3H -TdR and cells were collected 16 hours later. Incorporation of ^3H -TdR was measured by liquid scintillation spectroscopy. Results are expressed as mean c.p.m. $\pm$ s.e.m. We have checked AC50 for its reactivity with other helper clones from the same individual but with different fine specificities. It did not react with 2 other helper clones (HA1.7, HA1.9) as well as other autologous clones (37, 53, 72).

and TLC6 could not stimulate CTLL to proliferate (31 c.p.m.) in contrast to TCGF (6,503 c.p.m.), as shown in Table 3. There was no effect of AC50 and TLC6 even in the presence of E⁻ cells.

Thus the induction of thymidine incorporation in TLC6 by AC50 was not due to back stimulation, and thus the bidirectional nature of the TLC6-AC50 interaction and its specificity for TLC6, but no other autologous clones would appear to indicate that AC50 recognizes antigen-specific receptors on TLC6. Because the reaction is autologous, and specific for TLC6 it cannot be directed against MHC or differentiation antigens. The criteria we have used to define the anti-idiotypic nature of clone AC50 (that is, specificity, stimulatory effects)

are the same as used for anti-idiotypic antisera (discussed in refs 10, 12).

The effects of clone AC50 on the helper response of TLC6 were assessed *in vitro* (Table 4). Very small numbers of AC50 diminished the response. Only 10 clone AC50 cells inhibited the helper response of 500 TLC 6 (to 100,000 E⁻ cells in a 200 μl culture volume) by 95%. Fifty cells of AC50 per culture abolished the response altogether. These results indicate that AC50 is a potent suppressor cell. The mechanism of suppression is unknown, but it is unlikely, due to the ratio of cells used (1 suppressor to 50 helpers) and the capacity of irradiated AC50 to induce TLC6 to incorporate thymidine (Table 2), that AC50

Table 2 Anti-idiotypic activity of clone AC50

Clone no.	Antigen				TCGF
	(5×10^2)	Clone AC50* (1×10^3)	(5×10^3)	Medium	
6	820 ± 124	1,160 ± 169	2,274 ± 322	17 ± 4	3,472 ± 313
6 + E ⁻ *	3,413 ± 369	4,074 ± 254	5,328 ± 590	31 ± 16	
6 + E ⁻ + Flu	10,311 ± 2,404	8,902 ± 866	7,386 ± 1,063	9,360 ± 495	
37	43 ± 9	59 ± 8	65 ± 16	35 ± 3	8,551 ± 1137
37 + E ⁻	38 ± 8	71 ± 18	32 ± 4	53 ± 13	
37 + E ⁻ + Flu	13,569 ± 1,861	9,978 ± 1,348	11,703 ± 1,018	14,325 ± 2,543	
53	27 ± 4	35 ± 12	17 ± 6	25 ± 6	6,217 ± 491
53 + E ⁻	33 ± 5	18 ± 2	24 ± 2	12 ± 1	
53 + E ⁻ + Flu	4,261 ± 434	4,954 ± 129	6,203 ± 631	4,829 ± 686	
72	23 ± 1	8 ± 2	19 ± 5	20 ± 5	4,142 ± 474
72 + E ⁻	41 ± 14	26 ± 3	13 ± 2	29 ± 3	
72 + E ⁻ + Flu	7,443 ± 1035	7,248 ± 131	8,130 ± 1076	7,055 ± 360	
HA1.7	17 ± 2	15 ± 2	10 ± 3	31 ± 6	3,447 ± 122
HA1.7 + E ⁻	33 ± 1	43 ± 8	18 ± 2	12 ± 1	
HA1.7 + E ⁻ + Flu	3,625 ± 329	4,288 ± 372	4,100 ± 105	3,814 ± 261	
CTLL	58	49 ± 11	115 ± 17	73 ± 14	10,361 ± 1,895
CTLL + E ⁻	101	146 ± 23	742 ± 109	60 ± 8	
CTLL + E ⁻ + Flu	20,183	17,081 ± 3,036	15,622 ± 1,388	18,762 ± 2,573	

Various numbers of irradiated AC50 cells (5×10^2 , 1×10^3 and 5×10^3 per well) were cultured with a panel of influenza A virus-specific TLCs (5×10^3 per well) in the presence or absence of autologous irradiated sheep erythrocyte (SRBC) rosette negative (E⁻) cells (5×10^3 per well) and Influenza A virus. E⁻ cells were separated from those cells forming rosettes with AET (aminoethylisoethioureonium bromide hydrobromide)-treated SRBC by centrifugation over Percoll (1,080 g ml⁻¹). Proliferative responses were determined by the incorporation of ^3H TdR as described in the legend to Table 1. Results are expressed as mean c.p.m. ($\pm$ s.e.m.). CTLL, influenza specific T lymphocyte line.

* E⁻ and AC50 irradiated 2,500 rad.

Table 3 Lack of back stimulation in mixed cultures of AC50 and TLC6

Responder cells	Stimulation (c.p.m. $\pm$ s.e.m.)					TCGF
	AC50	6	AC50+6	FluA	Medium	
CTL	43	17	31	15	23	6,503
	± 7	± 4	± 9	± 2	± 6	± 294
CTL+E ⁻	12	21	57	14,689	27	—
	± 5	± 5	± 8	$\pm 2,229$	± 4	

Irradiated AC50 (5×10^3 cells per well) and TLC6 (5×10^3 cells per well) were cultured alone or together with an influenza specific cultured T lymphocyte line (not the same as in Table 2) CTL (5×10^3 cells per well) in the presence or absence of autologous irradiated E⁻ cells. As controls CTL cells were cultured with and without irradiated E⁻ cells and influenza A virus. Proliferative responses were measured and results expressed as described in Table 1.

is cytotoxic for TLC6. The mechanism of suppression remains to be elucidated, but because it occurs at the clonal level should be amenable to precise analysis.

These results indicate that as in the mouse, it is possible to generate a human specific suppressor T cell clone dependent on IL-2. In contrast to the mouse suppressor cell clone, specific for antigen, the specificity of the human suppressor clone was for the immunizing cell (and not other autologous helpers). The reciprocal stimulation, in the absence of back stimulation, suggests that the target of AC50 is the receptor for antigen of TLC6. This is a direct experimental test of the 'idiotype network' at the T cell level. The potency of the suppressor effect suggests that this type of suppressor cell may be of major importance in immune regulation.

The generation of suppressor T cells recognizing helper cell receptors resembles work reported by Binz and Wigzell, who injected antigen-specific lymphoblasts into autologous mice¹³. Specific unresponsiveness resulted, which was mediated by either T killer, suppressor or helper cells, or B cells¹³. Conceptually the results are analogous, but the experiments reported here were performed entirely *in vitro*, and at the clonal rather than the population level. There are reports of suppressor cell

pathways to the haptens nitrophenyl (NP) or azobenzene arsonate (ABA) for delayed hypersensitivity which involve second order suppressor cells, which recognize idiotype bearing cells¹⁴. However, the final suppression was not idiotype specific¹⁵. These experiments were also performed at the population level.

By possessing a clone of suppressor cells (AC50), specific for a clone of helper cells (TLC6), we have a powerful system for analysing the mechanism of T cell suppression. There is no reason to believe that the principle demonstrated here will not apply with other clones. Thus, situations with undesirable clones of helper (or killer) cells such as autoimmunity or perhaps even leukaemia may be controllable by suppressor cells of this type, generated *in vitro*.

This work was supported by US Office Naval Research Contract N000-14-77-C-0748 USPHS, and the Imperial Cancer Research Fund. We thank Drs Woody, Eckels and Lake for their helpful assistance in this work, Philippa Wells for typing the manuscript and Dr M. C. Raff for his helpful comments on the manuscript.

Received 30 July; accepted 28 September 1982.

1. Miller, G. (ed.) *T Cell Clones*, Immun. Rev. **54** (Munksgaard, 1981).
2. Fresno, M., Nabel, G., McVay Boudreau, L., Firthmayer, H. & Cantor, H. *J. exp. Med.* **153**, 1246-1259 (1981).
3. Lamb, J. R. et al. *J. Immun.* **128**, 233-238 (1982).
4. Jerne, N. K. *Ann. Immun. (Inst. Pasteur)* **125C**, 373-389 (1974).
5. Lamb, J. R., Woody, J. N., Hartzman, R. J. & Eckels, D. D. *J. Immun.* **129**, 1465-1470 (1982).
6. Eckels, D. D. et al. *Hum. Immun.* **4**, 313-324 (1982).
7. Callard, R. E. *Nature* **282**, 734-736 (1979).
8. Zanders, E. D., Smith, C. M. & Callard, R. E. *J. immun. Meth.* **47**, 333-339 (1982).
9. Fischer, A., Beverley, P. C. L. & Feldmann, M. *Nature* **294**, 166-168 (1981).
10. Infante, A. J., Infante, P. D., Gillis, S. & Fathman, C. G. *J. exp. Med.* **155**, 1100-1107 (1982).
11. Inouye, H., Hank, J. A., Alter, B. J. & Bach, F. H. *Scand. J. Immun.* **12**, 149-154 (1980).
12. Janeway, C. A., Sercarz, E. & Wigzell, H. (eds) *Immunoglobulin Idiotypes* (Academic, New York, 1981).
13. Binz, H. & Wigzell, H. *J. exp. Med.* **147**, 63-76 (1978).
14. Sy, M. S., Dietz, H., Germain, R. N., Benacerraf, B. & Greene, M. J. *exp. Med.* **151**, 1183-1195.
15. Sy, M. S., Nisonoff, A., Germain, R. N., Benacerraf, B. & Greene, M. J. *exp. Med.* **153**, 1415-1425 (1981).

Table 4 Cloned auto (anti-idiotypic) suppressor T lymphocytes

Helper T cells	Stimulus		Suppression (AC50 per ml)	Response Anti-influenza A virus antibody (ng ml ⁻¹)
	Autol. E ⁻ (5×10^5 per ml)	A/Texas/1/77 (0.5 HAU per ml)		
E ⁺ *	—	—	—	0
E ⁺	—	+	—	0
—	+	—	—	0
—	+	+	—	4 $\pm$ 1
E ⁺	+	+	—	145 $\pm$ 26
TLC6†	—	—	—	0
+	+	—	—	0
+	+	+	—	208 $\pm$ 22
+	+	+	50	10 $\pm$ 2
+	+	+	250	3 $\pm$ 1
+	+	+	500	0

Cloned helper T cells (TLC6; 2.5×10^3 per ml) or E⁺ cells (5×10^5 per ml) were cultured with autologous E⁻ cells in the presence of influenza A virus (A/Texas/1/77; 0.5 HAU ml⁻¹) in RPMI 1640 supplemented with 10% horse serum in 96 well round bottom microtitre trays. TLC AC50 cells were added at 50, 250 and 500 cells per ml at the initiation of the cocultures each of 200 μ l. After 6 days incubation at 37 °C, triplicate cultures were washed and recultured in 150 μ l of RPMI 1640 supplemented with 5% fetal calf serum (FCS). Supernatants were collected after 24 h and assayed for anti-A/Texas/1/77 antibody using a solid phase enzyme immunoassay. Antibody production was determined by reading absorbance at 405 nm of the transformed substrate of alkaline phosphatase⁹. The actual amount of antibody (ng ml⁻¹) $\pm$ standard error of triplicate cultures. Results are expressed as the mean (ng ml⁻¹) and for each triplicate the s.e. was <20%. Background responses of TLC6, E⁺ and E⁻ cultured alone or with influenza A virus were measured. * 5×10^5 ml⁻¹; † 2.5×10^3 ml⁻¹.

A protein with kinase and phosphatase activities involved in regulation of tricarboxylic acid cycle

D. C. LaPorte & D. E. Koshland Jr

Department of Biochemistry, University of California, Berkeley, California 94720, USA

Phosphorylation-dephosphorylation reactions as a form of regulatory control are assuming increasing importance in our understanding of cellular processes. The phosphorylation-dephosphorylation reaction was the first example of covalent modification in control and has been found to be present in almost every type of cellular process including synthesis of cell constituents, metabolic pathways and even gene expression¹. It has assumed increasing importance as a result of the finding that some proteins coded for by carcinogenic viruses apparently are involved in a phosphorylation reaction^{2,3}. To better understand the principles underlying these processes, we have studied the role of phosphorylation of isocitrate dehydrogenase (IDH) in the regulation of the branch point between the tricarboxylic acid (TCA) cycle and the glyoxylate bypass. We report here that in the process of purifying the kinase and phosphatase, we made two surprising observations—the kinase and phosphatase activities appear to be associated with the same protein, and the phosphatase activity has an absolute requirement for ATP.

During growth on acetate, *Escherichia coli* achieves net synthesis of cellular intermediates by induction of the glyoxylate bypass⁴. This anapleurotic pathway competes with the TCA cycle, the principal producer of metabolic energy, for a limiting amount of isocitrate. The branch point between these pathways

has been shown to be regulated, at least in part, by phosphorylation of IDH⁵⁻⁷.

Our first goal in the study of this branch point was to purify the relevant enzymes. We purified IDH kinase by using three steps: ammonium sulphate fractionation, chromatography on DEAE-Sephacel and affinity chromatography using IDH immobilized on Sepharose 4B. Absorption of the IDH kinase to the affinity column was found to require the non-hydrolysable ATP analogue, 5'-adenylylimidodiphosphate. After sample application, the column was washed with buffer containing the analogue until the effluent was free of protein. The bound enzyme was then eluted by washing the column with a buffer which did not contain the analogue. Figure 1 presents the results of electrophoretic analyses of the peak activity fractions; they each contain a single protein band, the intensity of which was proportional to the amount of IDH kinase activity. The mobility of the protein band indicates a molecular weight (MW) for the protein of 66,000.

We then proceeded to purify the IDH phosphatase. During preliminary experiments it became apparent that the steps needed for phosphatase purification paralleled those for the kinase and, indeed, that the ratios of the two activities remained almost constant. We therefore focused our attention on attempts to separate the two activities by various techniques. In our first attempt we used ion-exchange chromatography on DEAE-Sephacel (Fig. 2). This column yields good resolution of the applied protein, typically producing 8–12-fold purification of the IDH kinase. Despite this efficiency, the kinase and phosphatase activities eluted superimposably. Our next attempt to separate the two activities used a Sephadex G150 column, and again the activity peaks were indistinguishable. Finally, we re-examined our IDH-Sepharose affinity column and found that both the kinase and the phosphatase absorbed to the column in the presence of the ATP analogue and were superimposably eluted when the analogue was

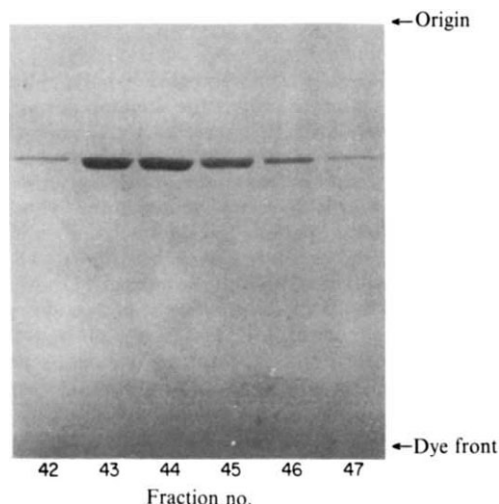


Fig. 1 SDS-polyacrylamide gel electrophoresis of purified IDH kinase following IDH-Sepharose chromatography. Partially purified IDH kinase was prepared by ammonium sulphate precipitation and DEAE-Sephacel chromatography. The sample, in buffer A (25 mM MOPS pH 7.0, 50 mM NaCl, 5 mM MgCl₂, 2 mM 2-mercaptoethanol, 0.5 mM 5'-adenylylimidodiphosphate), was applied to an affinity column prepared by immobilization of IDH on Sepharose 4B. The column was then washed with buffer A until the effluent was free of protein. The bound IDH kinase was then eluted by washing the column with 25 mM MOPS pH 7.0, 50 mM NaCl, 5 mM EDTA, 2 mM 2-mercaptoethanol. After assay for IDH kinase, samples from peak activity fractions were subjected to SDS-gel electrophoresis on 10% polyacrylamide gels. The gels were stained with Coomassie brilliant blue and destained with 10% acetic acid. Molecular weight standards, electrophoresed simultaneously, were phosphorylase b (94,000), bovine serum albumin, BSA (67,000), ovalbumin (43,000) and carbonic anhydrase (30,000).

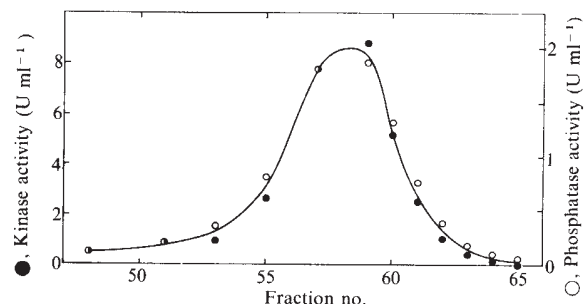


Fig. 2 Chromatography of IDH kinase and IDH phosphatase on DEAE-Sephacel. Following cellular disruption, centrifugation and ammonium sulphate precipitation, the sample was dialysed against buffer B (25 mM MOPS pH 7.0, 100 mM NaCl, 1 mM EDTA, 2 mM 2-mercaptoethanol), then applied to a DEAE-Sephacel column (2.5 × 35 cm) equilibrated in buffer B. After application, the column was washed with 100 ml buffer B and eluted with a linear gradient from 100 to 300 mM NaCl (750 ml each). Fraction size was 10 ml, flow rate 150 ml h⁻¹. IDH kinase was assayed by its ability to inhibit IDH in an ATP-dependent manner. The reaction mixture contained buffer C (25 mM MOPS pH 7.5, 5 mM MgCl₂, 100 mM NaCl, 2 mM dithiothreitol, 1 mM EDTA, 1 mM ATP, 0.5 mg ml⁻¹ BSA and 0.6 μM IDH in a final volume of 50 μl. The reaction was started by adding ATP and terminated by addition of 50 μl 20 mM EDTA. Aliquots were then withdrawn and assayed for IDH activity in a reaction mixture containing 25 mM MOPS pH 7.5, 250 μM DL-isocitrate, 100 μM NADP, 5 mM NaCl. One unit of IDH kinase will produce 50% inhibition of IDH in 1 min. Calculations consider the IDH kinase reaction to be pseudo-first order with respect to IDH. IDH-phosphatase was assayed by its ability to catalyse the release of ³²P from phospho-IDH. The reaction mixture contained buffer C, 1 mM ATP, 2 mg ml⁻¹ BSA and 0.1 μM ³²P-phospho-IDH (~20,000 c.p.m.) in 100 μl. The reaction was begun by addition of phospho-IDH and stopped by adding 150 μl of 20% trichloroacetic acid. The precipitated protein was removed by centrifugation and aliquots of the supernatant withdrawn for determination of ³²P. One unit of IDH phosphatase will catalyse the release of 50% of the ³²P in 1 min. Calculations consider the IDH phosphatase reaction to be pseudo-first order with respect to phospho-IDH. Typically, less than 20% of the phospho-IDH was hydrolysed during the reaction.

removed. The single protein band which eluted with the enzyme activity, as described above, showed a constant ratio of kinase to phosphatase activity. Thus, IDH kinase and IDH phosphatase are chromatographically indistinguishable when tested by ion-exchange, gel filtration and affinity column techniques. The purified activities exhibit a single protein band. Therefore, these activities appear to be associated with the same protein and possibly the same polypeptide chain. These observations do not, however, completely exclude the possibility that the two activities exist on highly similar proteins, or on non-identical subunits of a heterodimer.

Initial studies of the IDH phosphatase showed that the activity in crude extracts was too low to account for observed rates of dephosphorylation *in vivo* and was very unstable as regards purification attempts. In resolving these problems, we discovered that the phosphatase activity had an absolute requirement for ATP or ADP. The activator has actually been shown to be Mg-ATP and all studies have been performed in 5 mM Mg²⁺. This observation, together with the co-purification of kinase and phosphate activities, suggested that the activity identified as IDH phosphatase was, conceivably, the back reaction of the kinase. In such a case, ADP should be a far superior substrate as the ATP-dependent activity would actually result from the presence of ADP, either as a contaminant of the ATP or generated from it during the assay. In fact, ATP was found to be the superior substrate, with Michaelis-Menten constants of 100 and 400 μM for ATP and ADP, respectively.

To add weight to the conclusion that the ATP-dependent release of phosphate from phospho-IDH did not result from the ADP-dependent back reaction of the kinase, we tested the

effect of adding an ATP regenerating system (Table 1). As such a system efficiently converts ADP to ATP, it would be highly inhibitory to any ADP-dependent reactions. The ATP regenerating system had no effect on the ATP-dependent phosphatase activity. When the ADP-dependent activity was examined, conversion of ADP to ATP by the regenerating system actually increased the phosphatase activity to a level slightly greater than that observed with an equivalent concentration of ATP; this increase appears to be due to an unidentified minor contaminant in the ADP preparation.

The products of the kinase back reaction are ATP and IDH. Thus, if the kinase back reaction were responsible for the release of ^{32}P from ^{32}P -phospho-IDH, the label should appear in ATP. When the products of the ATP- or ADP-dependent phosphatase reactions were treated with activated charcoal, we found that none of the released ^{32}P was absorbed, although the nucleotide was absorbed quantitatively. This observation, taken together with the results given above, indicates that the ATP-dependent phosphatase reaction is not the back reaction of the kinase.

The location of IDH kinase and phosphatase on the same protein resembles two previously described systems. It has been shown that the enzymes responsible for the adenylation-deadenylation of glutamine synthase in *E. coli* are located on the same polypeptide chain^{8,9}. In addition, it has been reported recently that the rat liver 6-phosphofructo-2-kinase and fructose-2,6-bisphosphatase activities, which catalyse synthesis and breakdown of fructose-2,6-bisphosphate, are also physically associated^{10,11}. It is interesting to note that, in addition to representing examples of the physical association of opposing activities, all three enzyme systems have major roles in the regulation of intermediary metabolism.

The possible advantages of localizing opposing activities on the same protein presents an intriguing problem. The occurrence of at least three such systems suggests that this is not an evolutionary accident. Such a physical association might have the advantage of allowing both activities to be controlled from a single site on the protein. While such a simultaneous regulation may occur, this is not the only mechanism by which such effects can be achieved. Simultaneous regulation could occur as a result of the interaction of a ligand with two distinct sites, one associated with each activity. Alternatively, in cases where the interconverted molecule is a protein, effectors which bind to that protein could result in simultaneous regulation. Nevertheless, physical association of opposing activities might represent a favourable mechanism for simultaneous regulation.

In addition to the potential for simultaneous regulation, the physical association of kinase-phosphatase for IDH may have been important in the evolution of the protein. The phosphatase site may have arisen from the kinase site by an imperfect gene duplication, possibly including a fusion event. If the imperfect duplication of the kinase site produced a phosphatase activity, the new site would be already designed to act on the common substrate. Although speculative, this suggestion explains the ATP requirement with the nucleotide achieving the same induced conformational change as for the kinase reaction, but with the specificity of the acceptor function altered to allow access by water.

Table 1 Effect of an ATP regenerating system on IDH phosphatase activity

Nucleotide		Phosphatase activity (mU)
ATP (0.1 mM)	No addition	12.4
	+ ATP regenerating system	11.9
ADP (0.1 mM)	No addition	4.4
	+ ATP regenerating system	14.0

Phosphatase assays were performed as described in Fig. 2 legend. Activity is presented in milliunits (mU). The ATP regenerating system consisted of 2 mM phosphocreatine and 10 U ml⁻¹ creatine phosphokinase.

We thank Kathy Parker for technical assistance and Ken Walsh for helpful discussions. D.C.L. was supported by a NIH postdoctoral fellowship. This research was supported by grants from the NIH and the NSF.

Received 2 August; accepted 13 October 1982.

1. Krebs, E. G. & Beavo, J. A. *Rev. Biochem. Sci.* **48**, 923-959 (1979).
2. Collett, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021-2024 (1978).
3. Levinson, A. D., Oppermann, H., Levintow, L., Varmus, H. E. & Bishop, J. B. *Cell* **15**, 561-572 (1978).
4. Kornberg, H. L. & Madsen, N. B. *Biochim. biophys. Acta* **24**, 651-653 (1957).
5. Bennett, P. M. & Holms, W. H. *J. gen. Microbiol.* **87**, 37-51 (1975).
6. Garnak, M. & Reeves, H. C. *Science* **203**, 1111-1112 (1979).
7. Wang, J. Y. J. & Koshland, D. E. Jr *J. biol. Chem.* **256**, 4640-4648 (1981).
8. Caban, C. E. & Ginsburg, A. *Biochemistry* **15**, 1569-1580 (1976).
9. Rhee, S. G., Park, R., Chock, P. B. & Stadtman, E. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3138-3142 (1978).
10. El-Maghrabi, M. R., Claus, T. H., Pilakis, J., Fox, E. & Pilakis, S. J. *J. biol. Chem.* **257**, 7603-7607 (1982).
11. El-Maghrabi, M. R., Fox, E., Pilakis, J. & Pilakis, S. J. *Biochem. biophys. Res. Commun.* **106**, 794-802 (1982).

A second plasma calcium-lowering peptide from the human calcitonin precursor

Iain MacIntyre, Carmel J. Hillyard, Patricia K. Murphy, John J. Reynolds*, Rose E. Gaines Das† & Roger K. Craig‡

Endocrine Unit, Royal Postgraduate Medical School, Du Cane Road, London W12 0HS, UK

* Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, UK

† National Institute for Biological Standards and Control, Holly Hill, Hampstead, London NW3 6RB, UK

‡ Courtauld Institute of Biochemistry, Middlesex Hospital Medical School, Mortimer Street, London W1P 7PN, UK

Calcitonin is a 3,500-molecular weight (MW) polypeptide whose main function in man is to protect the skeleton during growth, pregnancy and lactation¹. It is secreted by C-cells which are localized in the thyroid in mammals, but may be present mainly in the ultimobranchial body and lung of more primitive vertebrates². The hormone is also concentrated in the hypothalamus of several vertebrate species^{3,4}, including man⁵, and in the primitive brain of the chordate *Ciona intestinalis*^{6,7}. Based on cloning and subsequent analysis of recombinant DNA, recently we have reported the partial nucleotide sequence of the human calcitonin precursor mRNA^{8,9}. This study showed that calcitonin resides towards the C-terminal end of its precursor protein, flanked by cryptic peptides of unknown significance. We now report that the synthetic C-terminal flanking peptide (PDN-21) predicted from the nucleotide sequence is highly active biologically—it approaches the potency of calcitonin in reducing the level of plasma calcium. As immunological evidence suggests that PDN-21 is synthesized *in vivo* and circulates in man, it may represent a novel hormonal cleavage product from the human calcitonin precursor with a physiological role in Ca²⁺ regulation.

Increasing numbers of small peptide hormones have been shown to be synthesized as high molecular weight precursor proteins that are subsequently processed to yield one or more peptides of biological significance^{10,11}. Thus it was of interest to determine whether the predicted peptides flanking human calcitonin within the precursor possessed biological activity or were present only to ensure correct processing of the calcitonin precursor protein⁹.

To investigate this possibility, the human C-terminal flanking cryptic peptide (H-Asp-Met-Ser-Ser-Asp-Leu-Glu-Arg-Asp-His-Arg-Pro-His-Val-Ser-Met-Pro-Gln-Asn-Ala-Asn-OH) corresponding to the last 21 amino acid residues predicted from the coding sequence of the calcitonin precursor mRNA⁹, was

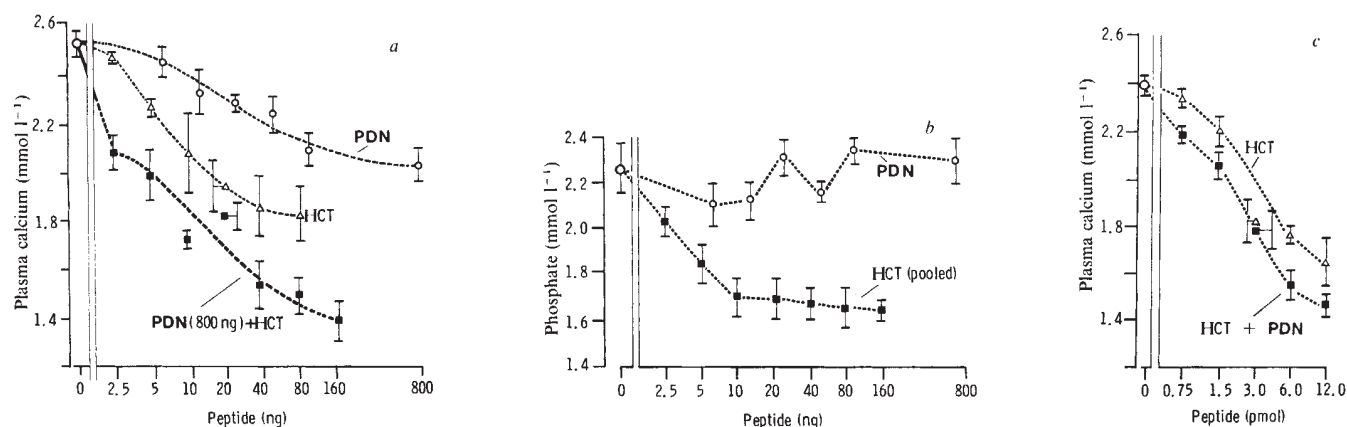


Fig. 1 Effect of PDN-21 on plasma calcium and phosphate. After preliminary experiments had shown that the time courses of the effects of PDN-21 and human calcitonin (HCT) on plasma calcium were similar, peptide was injected in sodium acetate buffer (pH 4.6) into the tail vein of female Wistar rats (mean weight 45 g, range 42–48 g; Oxfordshire Laboratory Animal Colonies) and blood taken from the aorta after 30 min in a randomized bioassay²³. *a*, Plasma calcium was measured colorimetrically by the method of Gindler and King²⁴ and the values confirmed by atomic absorption spectrophotometry. Each point is the mean $\pm$ s.e.m. from three rats. ○, PDN-21; Δ, human calcitonin (WHO 1st International reference preparation of calcitonin hormone for bioassay 70/234, supplied by the National Institute for Biological Standards and Control); ■, human calcitonin + PDN-21 (800 ng PDN-21 was injected with each dose of calcitonin). Analysis of variance showed highly significant ($P < 0.001$) dose-related plasma calcium differences among the three treatment groups. The dose giving 50% response (ED_{50}) was 7 and 24 ng for human calcitonin and PDN-21, respectively. Over the linear dose-response range, 1 ng of PDN-21 had the same effect as 0.17 ng human calcitonin; in the human calcitonin + PDN-21 group, 1 ng human calcitonin was as active as 5 ng given alone. Three previous unrandomized experiments gave essentially similar results. *b*, Plasma inorganic phosphate^{25,26} was measured in the same rats shown in *a*. PDN-21 produced no significant change at any dose level (○) but both the calcitonin and calcitonin + PDN-21 groups showed highly significant ($P < 0.001$, analysis of variance) dose-related reductions. These latter two groups did not differ and the results were therefore pooled (■). Values shown are mean $\pm$ s.e.m. *c*, In a separate batch of female Wistar rats (mean weight 51 g, range 51–53 g) from the same supplier, the effect of calcitonin alone (2.5–40 ng; Δ) was compared with that of calcitonin injected with equimolar amounts of PDN-21 (1.8–28 ng; ■). Each point is the mean for five rats $\pm$ s.e.m. The ED_{50} was 15 and 9 ng for human calcitonin alone and for calcitonin plus PDN-21, respectively. When equimolar PDN-21 was present, 1 ng of calcitonin was as active as 1.6 ng calcitonin alone (95% confidence limits 1.2–2.1 ng).

synthesized (Peninsula Laboratories) by solid-phase methodology, and the identity and homogeneity of the product established by amino acid analysis, high voltage electrophoresis and reverse-phase HPLC. We have called the peptide PDN-21 according to the terminology of Tatemoto and Mutt¹². This peptide was used to obtain preliminary evidence of biological activity and to raise the antisera required to establish its presence in human tissues and sera.

Initial pharmacological studies yielded striking results. PDN-21 is highly active in reducing plasma Ca^{2+} in the rat (Fig. 1*a*) although the maximum effect is less than that produced by calcitonin. Our findings also indicate that the potency of human calcitonin is increased by the presence of PDN-21; the enhancement was fivefold (Fig. 1*a*) when maximal doses were used but still 60% when equimolar PDN-21 was given with human calcitonin (Fig. 1*c*). As the two peptides are presumably secreted in equimolar amounts, these results imply that PDN-21 may be important in the regulation of plasma Ca^{2+} *in vivo*. Further, because the effect of PDN-21 on calcium is additive to that of calcitonin, it is possible that the peptide acts at a different receptor. This action of PDN-21 (800 ng) is still obvious in nephrectomized animals (Fig. 2) and is therefore unlikely to be due to loss of Ca^{2+} but rather to a redistribution, perhaps involving bone. This is consistent with results from 48-h mouse calvariae explants in culture (Table 1); there is clear evidence of a direct effect of PDN-21 on ^{45}Ca -release from the prelabelled bones ($P = 0.02$ at the higher dose). As yet, we have not determined whether PDN-21 has an important role in Ca^{2+} regulation and bone physiology. It is also intriguing to speculate whether the hypocalcaemia that follows hypercalcaemic thyroid perfusion¹³ is due solely to the effect of calcitonin release, or depends also on this new candidate hormone.

It is important to note that this effect on plasma calcium is highly specific. Thus, no effect was observed on plasma sodium, potassium, albumin or plasma phosphate (Fig. 1*b*). This latter finding clearly distinguishes PDN-21 from calcitonin, which lowers both plasma calcium and phosphate¹⁴. The glucose and magnesium levels did not differ significantly in relation to dose

although the latter were slightly higher overall. There were no effects in guinea pigs and rats on smooth muscle contraction (ileum and bronchus), urine flow, milk ejection pressure or blood pressure.

Our experiments thus establish the biological activity of PDN-21 and suggest it has a role as a calcium-lowering hormone in man. However, immunoassay with antiserum specific for

Table 1 Comparison of the effectiveness of human calcitonin and PDN-21 in inhibiting cell-mediated ^{45}Ca -release from prelabelled mouse calvariae *in vitro*

	% Inhibition of cell-mediated ^{45}Ca -release $\pm$ s.e.m.
PDN-21	
40 $\mu g\ ml^{-1}$	63.6 $\pm$ 14.3 (5)
10 $\mu g\ ml^{-1}$	25.5 $\pm$ 9.9 (4)
Human calcitonin	
0.2 $\mu g\ ml^{-1}$	62.8 $\pm$ 8.2 (5)
0.02 $\mu g\ ml^{-1}$	48.6 $\pm$ 10.5 (7)

The method for studying the response of mouse bones to calcitonin preparations has been described in detail²¹. Cell-mediated release of ^{45}Ca into the culture medium was calculated by subtracting the amount of exchangeable ^{45}Ca released into the medium by dead explants of 6-day-old prelabelled mice of the same litter from that released by the living treated explant²². The number of treated explants is shown in parentheses; the data were obtained after treatment of explants for 48 h *in vitro*. Although a higher dose of PDN-21 than human calcitonin was required, no attempt has been made to calculate 'potency' ratios because the dose-response curves for human calcitonin and PDN-21 are not parallel, and the relative stability of the peptides in the 48-h culture conditions is unknown. More work is necessary to establish the nature of the action of PDN-21 on explants and whether it is a true inhibition of resorption. Similar experiments with explants prelabelled with 3H -proline, which is incorporated into collagen, should be rewarding. In tests on uptake of ^{45}Ca into explants *in vitro* there did not seem to be any effect of PDN-21 over controls, therefore we conclude that it probably does not alter bone accretion.

PDN-21 shows that the peptide circulates in concentrations approximately equimolar with calcitonin in human medullary carcinoma of the thyroid (Fig. 3). As PDN-21 is also measurable in normal human thyroid tissue, the peptide presumably circulates in a healthy subject, suggesting a normal hormonal role. Although studies of the action of the equivalent rat calcitonin C-terminal peptide (CCP)^{15,16} have not been reported, measurements in the rat are in agreement with our results from immunoassay. Thus, antisera against synthetic CCP show that this peptide is present in equimolar amounts with calcitonin in normal rat thyroid tissue, rat medullary carcinoma of the thyroid (MCT), and is secreted by monolayer cultures of rat MCT cells^{17,18}.

Since the submission of this work, Amara and co-workers¹⁹ have described alternative, apparently tissue-specific RNA processing of rat calcitonin gene transcripts, resulting in the generation of mRNAs encoding two different polypeptide products: calcitonin, and a calcitonin gene-related peptide (CGRP). Calcitonin mRNA was the major species in the normal thyroid but CGRP mRNA predominated in the hypothalamus. The equivalent splicing mechanism giving rise to a human CGRP-like mRNA has yet to be described and it remains to be determined whether or not rat CGRP mRNA is translated into an active product in the brain.

Two main conclusions follow from our work. First, the human calcitonin precursor is a true polypeptide with at least two biologically active cleavage products and falls into the group of precursor proteins typified by pro-opiomelanocortin. Second, in addition to a possible involvement in Ca^{2+} regulation in health, PDN-21 may have therapeutic potential. Its intrinsic activity on calcium, together with its enhancement of the activity of calcitonin, suggests that it may be useful in the treatment of hypercalcaemia or of Paget's disease, where human calcitonin is at present the treatment of choice²⁰.

We thank Drs Sonia Jupp, John Stevenson and Elizabeth Hughes and Mr G. Abeyasekera for assistance. We also thank

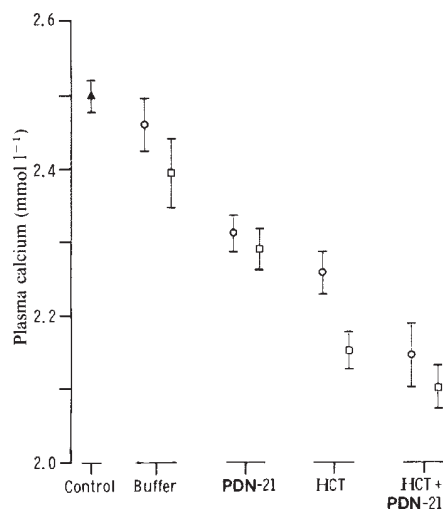


Fig. 2 The effect of nephrectomy on the calcium-lowering action of PDN-21 and calcitonin. Female Wistar rats (mean weight 56 g, range 57–58 g; Charles River, UK) were divided into groups of six, anaesthetized with ether and the kidneys exposed from a posterior incision (sham groups, □) or removed (○). Acetate buffer alone (0.2 ml), PDN-21 (800 ng), human calcitonin (40 ng) and both peptides together, were injected in buffer immediately after operation into both sham and nephrectomized groups. All the peptide preparations produced highly significant calcium reductions ($P < 0.005$). The action of PDN-21 was unaltered by nephrectomy although human calcitonin had a slightly smaller effect after removal of the kidneys ($0.01 < P < 0.05$). The control group (▲) was injected with acetate but not anaesthetized. Rats from this supplier are slightly less sensitive to both human calcitonin and PDN-21.

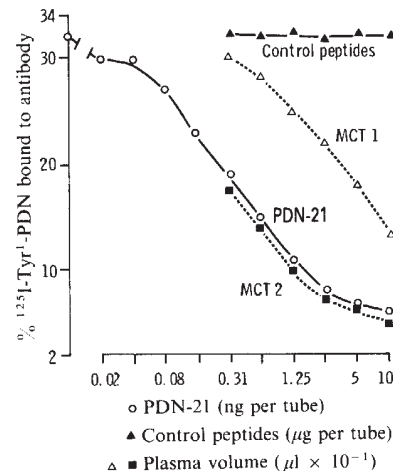


Fig. 3 Immunoassay of PDN-21 in human plasma. Tyr¹-PDN-21 (Peninsula Laboratories) was coupled to ovalbumin²⁷ and an anti-serum raised in rabbits. The serum was used at a dilution of 1:1,200 with ¹²⁵I-Tyr¹-PDN-21 as tracer. The presence of 10 µg per tube of luteinizing hormone-releasing hormone, salmon calcitonin, human calcitonin, somatostatin, ACTH₁₋₂₄ (Synacthen; Ciba-Geigy), porcine ACTH₁₋₃₉, human parathyroid hormone, oxytocin, insulin, pro-opiomelanocortin, human prolactin, or human growth hormone did not interfere in the assay (▲). The limit of sensitivity was ~40 pg per tube, the intra-assay variability < 5% and the inter-assay variation ± 10%. We studied plasma from 16 patients with MCT proven at operation; parallel displacement curves were found in 15 of the 16 cases. Plasma PDN-21 ranged from 0.6 to 138 ng ml⁻¹ and was approximately equimolar with human calcitonin (range 0.8–314 ng ml⁻¹). Results for plasma dilutions from two of the 16 MCT patients are shown together with a displacement curve for synthetic PDN-21 (○). Patient 1 (MCT 1, △): PDN-21 11 ng ml⁻¹, human calcitonin 18 ng ml⁻¹; patient 2 (MCT 2, ■): PDN-21 103 ng ml⁻¹, human calcitonin 314 ng ml⁻¹. As expected due to its equimolar relationship to normal calcitonin levels of 20–100 pg ml⁻¹, PDN-21 was not measurable in 12 normal subjects at this assay sensitivity.

Professor Gordon Bisset for pharmacological studies in guinea pigs and rats. Our experiments were supported in part by Celltech Ltd and the MRC.

Received 29 June; accepted 21 October 1982.

- Stevenson, J. C., Hillyard, C. J., MacIntyre, I., Cooper, H. & Whitehead, M. I. *Lancet* **ii**, 769–770 (1979).
- Ravazzola, M., Orci, L., Girgis, S. I., Galan, F. & MacIntyre, I. *Cell Biol. Int. Rep.* **5**, 937–944 (1981).
- Yui, R., Yamada, Y., Kayamori, R. & Fujita, T. *Biomed. Res.* **2**, 208–216 (1981).
- Galan, F., Rogers, R. M., Girgis, S. I. & MacIntyre, I. *Brain Res.* **212**, 59–66 (1981).
- Fischer, J. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7801–7805 (1981).
- Girgis, S. I. et al. *J. Endocr.* **87**, 375–382 (1980).
- Fritsch, H. A. R., Van Noorden, S. & Pearce, A. G. E. *Cell Tissue Res.* **202**, 263–274 (1979).
- Allison, J., Hall, L., MacIntyre, I. & Craig, R. K. *Biochem. J.* **199**, 725–731 (1981).
- Craig, R. K., Hall, L., Edbrooke, M. R., Allison, J. & MacIntyre, I. *Nature*, **295**, 345–347 (1982).
- Nakanishi, S. et al. *Nature* **278**, 423–427 (1979).
- Comb, M., Seeburg, P. H., Adelman, J., Eiden, L. & Herbert, E. *Nature* **295**, 663–666 (1982).
- Tatemoto, K. & Mutt, V. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6603–6607 (1981).
- Foster, G. V. et al. *Nature* **202**, 1303–1305 (1964).
- Gudmundsson, T. V., MacIntyre, I. & Soliman, H. A. *Proc. R. Soc. B* **164**, 460–477 (1966).
- Amara, S. G., David, D. N., Rosenfeld, M. G., Roos, B. A. & Evans, R. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4444–4448 (1980).
- Jacobs, J. W. et al. *Science* **213**, 457–459 (1981).
- Birnbaum, R. S., O'Neil, J. A., Muszynski, M., Aron, D. C. & Roos, B. A. *J. biol. Chem.* **257**, 241–244 (1982).
- Amara, S. G. et al. *J. biol. Chem.* **257**, 2129–2132 (1982).
- Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. M. *Nature* **298**, 240–244 (1982).
- MacIntyre, I., Evans, I. M. A., Hobitz, H. H. G., Joplin, G. F. & Stevenson, J. C. *Arthritis Rheum.* **23**, 1139–1147 (1980).
- Reynolds, J. J., Minkin, C. & Parsons, J. A. *Calcified Tissue Res.* **4**, 350–358 (1970).
- Reynolds, J. J. in *Organ Culture in Biomedical Research* (eds Balls, M. & Monnickendam, M.) 355–366 (Cambridge University Press, 1976).
- MacIntyre, I., Galante, L. S. & Hillyard, C. J. in *Handbuch der Inneren Medizin* Vol. 6, 1A (eds Kuhlencordt, F. & Bartelheimer, H.) 623–634 (Springer, Heidelberg, 1980).
- Gindler, R. M. & King, J. D. *Am. J. clin. Path.* **58**, 376–382 (1972).
- Lawrence, R. *Ann. clin. Biochem.* **11**, 234–237 (1974).
- Dryer, R. L., Tammes, A. R. & Routh, J. I. *J. biol. Chem.* **225**, 177–183 (1957).
- Reichlin, M. *Meth. Enzym.* **70**, 159–165 (1980).

BOOK REVIEWS

Men of science, men of letters

L. Pearce Williams

NATHAN and Ida Reingold, historian of science and mathematician respectively, have conceived the bold idea of presenting the history of science in America from the beginning of the twentieth century until the outbreak of the Second World War in Europe in terms of letters written by American men of science. The various sections devoted to the organization of science in America and the development of the various sciences are prefaced by short essays by the editors intended to introduce the *dramatis personae* and to sketch in the background against which the letters may be intelligently viewed and considered. The work, as a whole, is introduced by a short section whose main purpose is to define science in its American context and to indicate the place that science played in twentieth century America.

There is a notable virtue to this approach, and I should like to stress it at the outset of a review which will be generally critical. The reader is offered a splendid feast of hitherto unpublished and utterly fascinating primary sources. There are letters from almost all the "greats" of modern American science, containing important insights into their personalities, their concern for the public face of science and their own scientific work. Here are candid remarks on the organization and content of American science from William Henry Welch, Ira Remsen, George Ellery Hale, Charles W. Eliot, Irving Langmuir, Norbert Wiener, Jacques Loeb, Alfred H. Sturtevant, Albert Einstein and John von Neumann, to mention only the most famous, with replies from such non-American scientists as Ernest Rutherford, Wilhelm Ostwald and Bertrand Russell. As these names indicate, the sciences of astronomy, medicine, genetics and physics are well represented and readers will be able to discern what the major problems, both social and intellectual, were which concerned these men in the twentieth century. By publishing these letters, the Reingolds have done an invaluable service to historians of American science who might not have recognized the riches that were at their disposal, and by calling attention to the depositories in which these letters were found they have indicated rich veins for further exploration.

In this sense, this is an indispensable volume for all students of the modern American scene and I recommend it highly

Science in America: A Documentary History 1900-1939. Edited by Nathan Reingold and Ida H. Reingold. Pp.490. ISBN 0-226-70946-9. (University of Chicago Press: 1982.) \$49, £26.25.

to them. It should also be on the library shelf of every practising scientist over the age of 40 who will find in it echoes of his own education and scientific upbringing. Such persons will read it with an enjoyment that the non-scientist cannot share. This is, then, an important work but it is also deeply flawed and it is to its flaws that I must now turn.

In their brief preface, the Reingolds wrestle with a problem that has bedevilled historians of science now for a generation, namely the "internal" versus "external" perspectives on the history of science. The first states that science is an autonomous activity that develops by an internal logic of its own; the second insists that scientists reflect the general culture that surrounds them in their own scientific works. Non-historians of science may well be puzzled that serious scholars find much to argue about here since common sense indicates that both propositions are true. There *is* an internal logic to the sciences and scientists *do* reflect their society. The trick is to discover when each of these applies.

The Reingolds are perfectly aware of the fact that science has both a public and a private face, and they try to deal with this by means of a metaphor. Instead of speaking of a "core" of hard science, dear to the internalist, and an "outside" to which the externalist appeals, they assume that

there is no core — everything is part of an interface. And if the core with its weighty contents, has the aura of permanence, the interface is like a thin film, almost a soap bubble. Ideas, people, institutions of all sorts — not merely the scientific — exist in the interface. It is too much for such fragility, and tensions are created which produce additional film.

I simply do not understand what this metaphor means. How science exists in a soap bubble, or how tensions at the interface create more film escape me. Nor does the metaphor do much for the Reingolds except to permit them to insist that pure and applied science are both important, and that scientific institutions as well as private biases and prejudices go into the making of science. No one, again, will

argue with this conclusion which could have been presented without the metaphor.

The metaphor, however, does get the Reingolds into serious trouble. By defining science as part of an interface, they eliminate any distinction between the private and public face of science. This, then, permits them to publish a book containing documents drawn solely from the private sector and clearly intended by them to be a documentary history of science in America. I have often criticized historians of science whose works have included no references to private papers because they seemed to me to be seriously deficient. This work sins equally heavily on the opposite side. We read a number of letters by George Ellery Hale on various topics, but no reader of this volume would be able to describe Hale's contributions to modern astronomy in any detail. Similarly, the great physiologist, Jacques Loeb, is well represented by letters that, occasionally, refer to his epoch-making scientific work but not a single one of his important published articles is present.

The list of such omissions is as long as the list of letters and any reader who expects to gain an understanding of the course of scientific thought in America in the first third of the twentieth century may be bitterly disappointed by this book. Had it been entitled, *American Scientists: A Documentary History*, there would have been a much closer relationship between title and contents.

Having already praised the variety and interest of the letters printed by the Reingolds, I should also raise some questions about their editorial judgement in presenting them. The editors try to disarm critics by announcing that "the documents are edited with a decent minimum of editing". I should like to disagree. Important information is withheld and this vitiates some of the scholarship involved. One of the fundamental commandments of the editing of historical documents is to locate such documents in both time and place. Throughout the volume, place is systematically excluded. This leads, at best, to confusion as the reader tries to work out where the writer is so that he can understand local allusions. At worst, it makes parts of the letters unintelligible. Works such as this should not force their readers to guess where information is available. This is not a trivial

point — anyone wishing to use these letters in other works of scholarship will have to go to their sources to find out this information, thus forcing a duplication of effort that could easily have been avoided.

While I agree that footnotes to letters should be as short and concise as possible, it is important that they also provide the information necessary for the reader to understand what he is reading. In too many letters, there are names which are not identified. When Alfred Sturtevant urges Milislav Demerec "to try and get hold of Pink Jesus (= Bug Forbes)" (p.327), I suspect that few readers will understand the allusion. Professor Forbes was one of Cornell's legendary figures whose beard and ruddy complexion gave him his nickname and who certainly deserves a footnote. In other cases, a person is identified, if giving a first name, last name and dates of birth and death can be counted as identification — surely a brief description of a

person's place in science would have been helpful? The Reingolds also do not always provide essential information for their readers about scientific matters. In the same letter from Sturtevant and the one immediately preceding it, Sturtevant mentions *Oenothera lamarckiana*. How many readers will immediately recognize this as the plant used by Hugo de Vries in his studies of mutation? The Reingolds say nothing.

Although these are serious criticisms, I would not like to close on a sour note. This is not an adequate documentary history of science in America, as its title insists; but it is a fascinating slice of scientific Americana that should prove both stimulating and useful to a wide variety of readers. □

L. Pearce Williams is the John Stambaugh Professor of the History of Science at Cornell University. He is the author, among other works, of Michael Faraday, A Biography (Chapman & Hall, 1965).

Integrated enzymes

Athel Cornish-Bowden

Fundamentals of Enzymology. By N.C. Price and L. Stevens. Pp.454. Hbk ISBN 0-19-857175-5; pbk ISBN 0-19-857176-3. (Oxford University Press: 1982.) Hbk £25, \$49; pbk £12.50, \$24.95.

ENZYMOLGY has a role in nearly all aspects of research in biochemistry, and is of central importance in the whole subject. Partly because of this, it is commonly taught to undergraduate students as a series of disconnected topics. The people who study protein structure often have little in common with those concerned with the control of metabolic processes in the living organism. A third group of people study enzyme kinetics, and among these there sometimes seems to be an unbridgeable gulf between those who are interested in the steady state and those who are interested in transients. Yet these and other aspects of enzymology are all part of the same subject and it is clearly desirable to integrate them.

The present book is an attempt to describe enzymology as a whole, from the structures and properties of isolated enzymes to their roles as components of complex metabolic systems. It is written at a level where it can easily be understood by undergraduates, but it has enough detail to be of value to any research worker with an interest in enzymes. It is not the first book to attempt such a synthesis, but it is easily the most successful that I have seen, and it deserves to become a standard textbook of general enzymology.

The book is informative and accurate about the aspects of the subject that I know best, and so I feel inclined to believe what it tells me when it ventures into less well known territory, for example (in a final chapter on enzyme technology) about the origin of the holes in Swiss cheese. I found few faults, none of them serious. Although the authors are not themselves confused about the distinction between the Hill coefficient and the number of binding sites on a protein, they may well confuse their readers because they use the same symbol for both quantities on nearly consecutive pages.

Although monstrosities such as SGOT make fleeting appearances (inevitable, perhaps, in a chapter on clinical enzymology), and molecular weights are mentioned throughout, the authors have in general made commendable efforts to refer to enzymes by their recommended names and to use modern units and terminology. Even the clinical chapter is largely written in English, perhaps the first time this has ever been achieved. □

Athel Cornish-Bowden is a Lecturer in Biochemistry at the University of Birmingham, UK, and author of Basic Mathematics for Biochemists (Chapman & Hall/Methuen, 1981).

In terms of analytical chemistry

T.S. West

A Dictionary of Chromatography, 2nd Edn. By R.C. Denney. Pp.229. UK ISBN 0-333-31667-3; US ISBN 0-471-87477-9. (Macmillan Press/Wiley: 1982.) £15, \$39.95. *A Dictionary of Spectroscopy*, 2nd Edn. By R.C. Denney. Pp.205. UK ISBN 0-333-31670-3; US ISBN 0-471-87478-7. (Macmillan Press/Wiley: 1982.) £15, \$39.95.

THESE two second-edition dictionaries of chromatography and spectroscopy are attractive additions to the chemist's bookshelf, where they should serve as useful guides to those not fully expert in either technique. The author has successfully put together a reasonably comprehensive set of terms and definitions and has translated them into SI terminology in keeping with present-day trends. The terminology of the IUPAC "Orange Book" of Analytical Nomenclature has been adhered to fairly closely, and the author has steered a consistently sagacious path through the minefield of inconsistencies that abound in these as in most other areas of nomenclature.

I was, however, disappointed to find no mention of "limit of detection" in either book or "sensitivity" in the spectroscopy volume. Both volumes similarly define the term "signal-to-noise", but in neither is it said exactly what a signal is and the explanation of noise will not be very helpful to the puzzled novice, even in the chromatography volume. However, terms such as these are perhaps the most difficult to define and the author may have been wise to avoid or define them minimally. On the other hand, useful reference could

certainly have been made to other terms such as "time-resolved spectra" and "background", particularly for the newcomer to these techniques.

In both books the author has made excellent use of explanatory diagrams, 55 each for both chromatography and spectroscopy, and in general the definitions are clear and very much to the point, particularly those in the chromatography volume. Each book includes an extensive list of references for further reading, but a glossary of the terms defined, with the page numbers, would have been useful even though the two books are alphabetical lists from start to finish.

The chromatography volume gives the impression of being more comprehensive and perhaps more sure-footed than that on spectroscopy. However it may well be that the latter serves a more useful purpose in that it deals with some of the newer electron spectroscopies — such as Mössbauer and Auger spectroscopy — that are less well covered in other dictionaries which tend to concentrate exclusively on the more familiar molecular, atomic and electron spectroscopy techniques.

Each book is timely and excellent in conception, and in remarking that I would like to have seen such and such a term defined, I am basically expressing my satisfaction with the other definitions and wishing that there could have been more of the same. Both volumes deserve to sell well; they are especially recommended for student use. □

T.S. West is Director of the Macaulay Institute for Soil Research, Craigiebuckler, Aberdeen.

Scattering information from Graz

Henryk Eisenberg

Small Angle X-ray Scattering. Edited by O. Glatter and O. Kratky. Pp.515. ISBN 0-12-286280-5. (Academic: 1982.) £43.60, \$81.

FOR SOME years study of the scattering of radiation has been a favoured approach in the investigation of a variety of low molar mass, colloid and biological systems. There are three principal techniques: light scattering which, using an effective wavelength of a few thousand Angstroms, has been revolutionized by the advent of the laser; small angle X-ray scattering, which is restricted to low scattering angles because of the much lower wavelength (around one Angstrom); and neutron scattering, which has recently come to the fore and shows greater flexibility with respect to wavelength (from around one to a few tens of Angstroms). Although the three phenomena share the same basic theory, interesting complementarities result from the specific nature of each of the processes.

Recently, small angle X-ray scattering has received a strong impetus following the introduction of one- and two-dimensional position-sensitive detectors, computerized data acquisition and treatment, and the use of powerful synchrotron radiation sources. These developments have shortened experimental times to the subsecond range, allowing the study of dynamic phenomena. Long viewed with suspicion because of unwieldy and error-prone procedures, and a tendency towards overinterpretation of an essentially information-poor experimental technique, small angle X-ray scattering has now matured into a position of considerably higher respectability.

Although the editors have evidently made efforts to broaden the spectrum of contributors, the book under review represents to a large extent the individual and collective achievements of the Graz school of Kratky and Porod. Porod himself presents a comprehensive and lucid discussion of the general theory, which owes much to his efforts. Statistical fluctuations are mentioned, though not in sufficient depth to relate them to the analysis of multicomponent systems. Instrumental methods, and data treatment and interpretation are covered in detail in the various chapters, as well as contrast variation, a trade-mark procedure in neutron scattering. Although the underlying theory is identical for X-rays and neutrons, no equivalent exists to the contrast variation achieved by manipulation of hydrogen-deuterium compositions in the neutron technique.

Discussions of applications range widely over proteins, nucleic acids and nucleoproteins, lipoproteins and membranes, and natural and synthetic high polymers in the dissolved and solid state, including

evaluation of the Kratky-Porod persistent chain, inorganic substances, and aggregates and micellar structure of small molecules in solution.

This volume is a useful though somewhat uneven compilation. Thus, in parts, quantities in equations are not properly defined, some awkward expressions intrude, and references are missing or incomplete. Also it is irritating that different symbols are sometimes used for the same parameter, even in the same chapter.

Altogether, however, the book does provide a good source of information for evaluating both past achievements and future promise in an important field of study. For those wishing to come to grips with most of the theoretical and practical aspects involved it will prove instructive reading. □

Henryk Eisenberg is a Professor in the Polymer Department at the Weizmann Institute, Rehovot, Israel.

The right time for retroviruses

Neil Wilkie

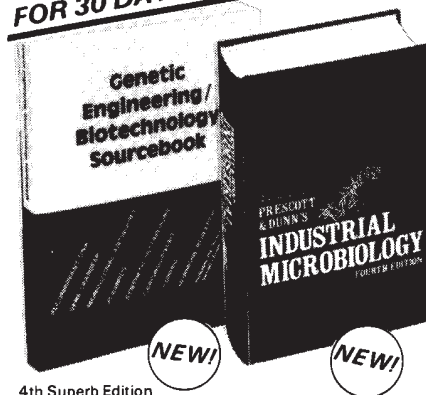
RNA Tumor Viruses. Molecular Biology of Tumor Viruses, Part 3, 2nd Edn. Edited by Robin A. Weiss *et al.* Pp.1,408. ISBN 0-87969-132-8. (Cold Spring Harbor Laboratory: 1982.) \$125 (US), \$150 (elsewhere).

ALL YOU ever wanted to know (and even more) about retroviruses, contained in one volume. That seems to be the promise of the new Cold Spring Harbor publication *RNA Tumor Viruses*, which together with the companion volume on DNA tumour viruses comprises the second edition of the popular series *Molecular Biology of Tumor Viruses*.

By and large, that promise is fulfilled by a notably clear and up-to-date review of a complex and fast-moving field. The decade elapsed since the publication of the first edition has witnessed remarkable progress in our understanding of the structure and function of retroviruses. It is not surprising, then, that the current volume is massive (it weighs 2 kg and runs to 1,400 pages) or that it is essentially a completely new book, rather than a revision of the first edition.

The stated aim of the book is to cater for both the novice and the expert, but novice and expert in what field? The study of

**SPECIAL
NATURE SUBSCRIBERS
OFFER-EXAMINE ON-APPROVAL
FOR 30 DAYS WITHOUT RISK**



4th Superb Edition
Prescott & Dunn's

Industrial Microbiology

Edited by Gerald Reed
Vice President, Amber Laboratories USA

The up-to-date guide and reference tool to industrial technology and applications in the microbiology of foods and food ingredients - with a special section on biological production of fuel alcohol.

Macmillan Hardcover 896 pp £42.50
ISBN 0 333 33630 5 1982

Genetic Engineering/ Biotechnology Sourcebook

Never before published! Profiles of 1,529 recent and ongoing genetic engineering/biotechnology research projects funded by 122 US government and non-profit organizations.

Covers research affecting: ★ Pharmaceuticals
★ Medicine and Diagnosis ★ Agriculture ★ Food Processing ★ Energy ★ Chemicals.

Reveals for each project ■ Nature and objectives of research ■ Funding source and money awarded
■ Names and addresses of individual researchers.

Special Yellow Pages reference section - tells you at a glance who's researching what, where, when and how.

Macmillan Softcover 350 pp £52.50
ISBN 0 333 34148 1 1982

Place your on-approval order today!

GLOBE
BOOK SERVICES LTD

Canada Road, Byfleet, Surrey KT14 7JL, England.
Tel: Byfleet (09323) 40397.

A MEMBER COMPANY OF MACMILLAN PUBLISHERS LTD.

EXAMINE ON 30 DAYS APPROVAL

To: Ian Jacobs, Globe Book Services Ltd., FREEPOST,
Canada Road, Byfleet, Surrey KT14 7BR, England.

(Please tick books required)

Please send me _____ copies of ☐ *Genetic Engineering/Biotechnology Sourcebook* at £52.50 each plus P&P (UK £2, Overseas £3) and/or _____ copies of ☐ *Industrial Microbiology* at £42.50 each, plus P&P (UK £2, Overseas £3) on 30 days approval. If I am not completely satisfied, I may return the volume(s) within that period for a full refund.

Tick preferred method of payment

☐ I enclose cheque for £ _____

☐ Charge my credit card £ _____

☐ Visa ☐ Access ☐ American Express ☐ Diners

Card No. _____ Expiry date _____

☐ Bill me. I need send no money now.

Signature _____
(Orders cannot be accepted without a signature)

Name _____ (PLEASE PRINT)

Address _____

Postcode _____

Allow 28 days for delivery
POST TODAY - NO STAMP NEEDED

Circle No. 67 on Reader Service Card.

retroviruses has provided insight into many important areas of biology removed from "straight" virology. Retrovirus genomes can be inherited as Mendelian genes as well as spread between and within species by infection; they mimic the effects of mobile genetic elements by mutation and alteration of the expression of chromosomal genes; they activate, recombine with and transduce cellular oncogenes; and they induce defects in the regulation of cell growth. The book is therefore of direct interest to virologists, molecular biologists, geneticists, cell biologists, tumour biologists and clinicians.

Does the book work? The text is divided into 11 chapters which deal firstly with molecular aspects of the genome, its replication and expression, and then with the biological aspects of expression, infection and pathogenesis. The self-contained chapter structure has resulted in considerable repetition and contributes to the overall length of the book. This is anticipated in the preface and on balance the strategy is acceptable; a little reiteration does no harm when the subject matter is so complex. The text has been researched in such formidable detail that complete novices may find some sections hard going. I imagine that undergraduates, for example, would need some guidance in the correct usage of the book. Nonetheless, the writing is generally clear and in most chapters there should be no difficulty in abstracting the information at different levels of complexity. Moreover, where doubt or ambiguity of interpretation is present, this is usually pointed out and the alternatives explained.

I particularly enjoyed the three chapters on the molecular biology of the retrovirus genome. They clearly explain the exquisite mechanisms by which the RNA genome is converted to a DNA copy in which a terminally redundant LTR sequence is created, how this is inserted into the chromosome and expressed using tactical variations on a strategy evolved to permit the synthesis of many gene products from a single genome sequence. The sections covering endogenous viruses also successfully deal with a complex subject while highlighting the evolutionary implications. Of course the pathogenicity, particularly the oncogenicity, of retroviruses is currently a centre of attention and readers will find the chapters on pathogenicity and oncogenes to be of

especial interest. These subjects are, in the main, well treated. Just enough (and no more) information on haematopoiesis is given to understand the concepts involved in virus-induced cellular disease. The several current models for the mechanistic interpretation of transformation are fairly represented and the review of oncogenes and their products is also well-balanced and up-to-date.

The final chapter on the identification of human RNA tumour viruses explains the dangers of overinterpreting initial results and introduces the only convincing isolate to date: the lymphoma-leukaemia virus associated with certain forms of adult T-cell leukaemia. A sting in the tail of the book is provided by a compilation of restriction endonuclease maps and primary nucleotide and amino acid sequences of some well-studied viruses.

Students of retrovirology and its practitioners will be not unaware of the clashes which often arise between some

dominant personalities with individual and exclusive viewpoints. Although only a small number of authors could be included in the writing of the current volume, my impression is that this has not been allowed to disturb the balance of the text and that the views expressed are reasonably representative of the field (not always the case in other volumes from the same stable).

I also sense that the timing of the book is critical: we are probably ending a phase in which details of the virus structure are the main concern, and beginning one in which the consequences of virus-cell interactions will predominate. This volume comfortably spans these areas. It is exhaustively researched and referenced, and will provide the standard reference text for the field for the foreseeable future.

Neil Wilkie is Head of the Tumour Virus Laboratory at the Beatson Institute for Cancer Research, Glasgow.

Looking at comets in Arizona

David W. Hughes

Comets. Edited by Laurel L. Wilkening. Pp.766. ISBN 0-8165-0769-4. (University of Arizona Press: 1982.) \$29.95.

TUCSON, Arizona was the place and March 1981 the time of the meeting of 175 scientists for the Sixty-first International Astronomical Union's Colloquium on "Comets: Gases, Ices, Grains and Plasma". The review papers presented at that meeting provide the basis for this book.

Comets have been newsworthy for centuries and scientists are forever using the latest observational techniques and theoretical gambits to delve into their mysteries. The 1980s are no exception. The impending return of Halley's comet in 1986 should see some visits by spacecraft; the successful operation of the International Ultraviolet Explorer has made H and OH observations in the extended coma a commonplace occurrence; the collection of cometary dust in the Earth's upper atmosphere is now routine; and the Pioneer Venus orbiter observations of the interactions of the solar wind with an atmosphere unprotected by a magnetic field gave a preview of what is to be expected in the cometary case.

"New theories and data are collected in this book as an introduction to comets for newcomers as well as a reference volume for specialists" — I quote from Laurel Wilkening's preface. Unfortunately I don't think a book can easily be both and this one certainly is not. It contains 29 papers grouped under headings such as Overview; Nucleus; Dust; Coma, Ion Tails and Solar Wind Interaction; Origin,

Evolution and Interrelations; and Basic Information and References. As with all conference proceedings the standard is mixed but there are some real gems — Delsemme on the chemical composition of cometary nuclei, Kresak on statistics and observational selection, Whipple on nuclei rotations, Sekanina on splitting, Wetherill and Revelle on the relationship between comets, meteorites and fireballs, A'Hearn on spectrophotometry, Feldman on the ultraviolet, Everhart on orbital evolution, Marsden and Roemer on astrometry and basic characteristics.

Such papers lift this conference proceedings from the category of "every library should purchase it" to "every comet researcher should have it on his personal bookshelf". If, however, critical comparisons must be made I do think the University of Arizona did a better job on the asteroids in 1979 than they have here on the comets. (For a review of *Asteroids*, edited by T. Gehrels, see *Nature* 287, 664; 1980.)

Also, when it comes to introducing a topic to newcomers it must be ensured that certain features are not omitted. This is especially difficult for the editor of conference proceedings who is restricted not only by the ease with which scientists raise travel money, and their predilections for Arizona in March, but also by the problems of convincing the participants to write up their papers in a suitable form. There are far too many gaps in this volume to make it a good introduction for the newcomer. □

David W. Hughes is a Lecturer in Astronomy and Physics at the University of Sheffield.

An English Lamarck

An English translation of Madeleine Barthélemy-Madaule's *Lamarck ou le Mythe du Précurseur* (Editions du Seuil, 1979) has just been published by MIT Press. The original edition was reviewed in *Nature* 286, 187; 1980.

The translation (by M.H. Shank) has the title *Lamarck the Mythical Precursor*, and costs \$17.50, £12.25.